Would the Permitting of Physician Assisted Suicide be a Desirable Extension of Patient Choice?

Andrew John Stanners

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

School of Philosophy, Religion and History of Science

Inter-Disciplinary Ethics Applied Centre

February, 2017

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

I would like to thank my family for their support and understanding while I was researching and working on this dissertation. I would also like to thank my supervisor, colleagues and friends for all their help.

Abstract

This dissertation argues that the permitting of Physician Assisted Suicide (PAS) is not a desirable extension of patient choice. Should PAS be made permissible by making it a live option, then certain patients may request it and be harmed by wrongful death. Furthermore, the harm to patients who suffer wrongful death as a result of requesting PAS trumps the harm to patients who must endure unbearable suffering should PAS not be permitted. The line of argument in defence of these claims is, first, that contrary to the common view, agents may sometimes be harmed when they are presented with an additional option. Second, the harm that may result from having an additional option occurs as a result of certain features of the agent or the context in which the agent is choosing. This second argument goes beyond previous ones because it explains two additional harms to an agent from a new option. These are harms resulting from three types of weak character and resulting from normative features of what I term the context of choice. Third, in order to decide whether or not to extend patient choice by permitting PAS, the harm to patients who may request it and suffer wrongful death and the harm to patients who are suffering unbearably and who cannot relieve their suffering through PAS must be weighed against one another. This weighing of harms is possible through insights gained from types of need. Categorical needs trump instrumental ones, and are also parallel to categorical harms. So, the categorical harm of wrongful death trumps lesser harms, such as suffering unbearably. Since the harm to patients who suffer wrongful death, should PAS be permitted, trumps the harm to other patients who are suffering unbearably, permitting PAS is not a desirable extension of patient choice.

Table of Contents

List of Tables

Introduction

Physician Assisted Suicide (PAS) is illegal in the United Kingdom[1] but there is currently an active public debate about whether it should be made permissible under the law[2], e.g. UK Government (2016). There are several strands of argument on either side of the debate, but the key arguments in favour of making PAS permissible centre on respecting patient autonomy and improving well-being, e.g. Brock (1992) and Dworkin, R. et al. (1997).The arguments against permissibility centre on potential harms to vulnerable people who may select or be pressurised or otherwise influenced into requesting PAS when this is not what they would want for themselves, e.g. Steinbock (2005) and Kamisar (1958). The literature on potential harms to vulnerable people, should PAS be made permissible, often focusses on empirical evidence, e.g. Battin et al. (2007). However, this approach has led to an impasse, since empirical approaches on either side of the debate have been construed as being subject to bias, e.g. Coggon (2006) and Dworkin, G. (2009). A further problem with empirical approaches to the debate is that they miss situations that I identify here in which patients change their minds and request PAS, when before PAS is made permissible they have a preference to stay alive, and would not choose PAS.

In this dissertation, I argue that the permitting of PAS is not a desirable extension of patient choice. I defend a claim that should PAS be made permissible, then some eligible patients who request it may potentially be harmed as a result of wrongful death: their deaths are wrongful either because PAS is not what they would have requested had it not been presented to them as a live option, or because their choice issues from a vice of character. Furthermore, I claim that the harm to patients who suffer wrongful death as a result of requesting PAS trumps the harm to selected patients who are suffering unbearably at the end of their lives, and who are denied PAS, should it not be made permissible. This is because the harm of wrongful death, and not the harm of unbearable suffering, derives its force from basic facts about human survival.

---

[1] Despite the Suicide Act 1961 stating that assisting a person to commit suicide is illegal, in the light of the non-prosecution of most cases of assisting suicide and on the application of Purdy, the Director of Public Prosecutions (Kier Starmer) has clarified conditions under which a person is more or less likely to be prosecuted if they assist someone to commit suicide (Director of Public Prosecutions, 2010).

[2] When I talk in this dissertation about making PAS permissible, I am talking about doing this in the legal sense.

Should PAS be made permissible, then it would become an additional option for patients who meet the relevant criteria: they may, then, either select PAS or not select PAS. It is commonly thought that agents stand to benefit when they are given an increased range of options, e.g. Mill (1982), Hurka (1987), and Reeve (1990). However, in chapter one, I defend a claim that an agent may potentially be harmed if the range of options available to her is increased. Specifically, I will argue that there are two potential harms that an agent may accrue if she is presented with an additional option, that are not identified in previous writings on harms from additional options by Dworkin, G. (1988) and Velleman (2015). The first of these two potential harms is made more likely if an agent has certain types of weak character, and the second may occur if she is adversely influenced by what I term "the context of choice". In line with my method in chapter one of analysing potential harms to an agent in the *general* case, where she is presented with an additional option, chapters two to five also consist in related arguments about the general case, rather than arguments specifically applicable to PAS. Having made progress with the general case in chapters one to five, I then apply these arguments when I return to the specific case of PAS as an additional option in chapter six.

In chapter two, in order to develop my line of argument that certain types of weak character may increase the chance that an agent is harmed should she be offered an additional option, I first explain a foundational account, from Aristotle (2002), of character. Second, I defend Aristotle's account of character against Kant's, and third, I compare Aristotle's account of character with contemporary accounts from Kupperman (1991), Goldie (2004), and Williams (1973), which I argue are not a significant advance on Aristotle. My defence of Aristotle against Kant is especially important since Aristotle affords a special role for emotion in character that is denied by Kant. Fourth, I also defend Aristotle's account of character from objections that have been raised from the perspective of situationist ethics, which claims that it is situational features alone that settle how an agent behaves, rather than character, e.g. Harman, G. (2000) and Doris (2002). So, my overall aim in chapter two until this point is to make Aristotle's account of character plausible, before responding to some alternative accounts and an objection to character. In the remainder of the chapter I defend a central theoretical role for the concept of character in an account of good ethical judgement. This is also important to my thesis because I am relying on the notion of weak character to explain

potential harms to an agent and, if character does not have a central role in good ethical judgement, then types of weak character cannot explain these harms.

Next, in chapter three, I defend a claim that selected features of an agent's character may make her more susceptible to harms from being given an additional option. In particular, I defend an account of three types of weak character—weakness of will (*acrasia*) (Aristotle, 2002), undue self-deprecation—a vice which is one of the defects of character associated with the virtue of humility (Richards, 1988)—and undue lack of confidence in judgements, which is also a vice. These types of weak character may increase the chance of an agent being harmed when she is presented with an additional option.

Alongside weak character, the second concept I appeal to, in order to explain potential harms that an agent may accrue when given an additional option, is what I term the "context of choice". In chapter four, then, in order to defend my account of the context of choice, I begin by noting a linguistic cue to be found in the etymology of the word context, that helps to identify features of a choice situation that are relevant for my purposes, namely features that are woven into an agent's decision-making and which may, thereby, influence this. Second, I use resources from literature on the theories of bounded rationality (BR) and adaptive preferences (AP) to delimit the context of choice to features of a choice situation that are either salient to an agent or should or should not be salient to her. Third, I illuminate the context of choice with a medical analogy. The medical analogy compares a context in which a doctor is diagnosing with a context in which an agent is choosing. This analogy helps to draw out important normative aspects of the context of choice, that explain how an agent may be harmfully influenced in a choice situation. She may be harmed if she either finds features of the choice situation salient that she should not find salient, or if she does not find features of the choice situation salient that she should find salient, in order for her to make a rational choice.

Clearly an agent may also *benefit* from having an additional option, so she may be harmed if she does not have this option. If a decision is being made about whether or not to offer an additional option (such as PAS), then the harms to one group of people if it is offered have to be weighed against the harms to the other group if it is not offered, in order that one side or the other may be given priority. So next, in chapter five, I defend a claim that certain harms can be prioritised. As a result of this prioritisation, some harms may trump other harms. The line of my argument about trumping harms,

first, does not derive from the literature on how harms may be weighed, e.g. in Taurek (1977) cases, since I claim that arguments about how to resolve these types of case, or their variants, have not been decisively resolved. Second, the literature on harms themselves is not useful for my purposes, since it does not aim to identify trumping harms. Nor, thirdly, does the notion of "rights as trumps" (Dworkin, R., 1989), offer significant help. My line of argument about trumping harms, then, is based on insights gained from arguments about needs from Megone (1992) and Wiggins (1998): I argue that harms that are parallel to so-called categorical needs are trumping ones and should, thereby, always take priority over non-trumping harms.

In chapters one to five, then, I argue that an agent may potentially be harmed when she is offered an additional option if she has certain types of weak character or the context of choice has a particular configuration. Furthermore, I argue that some harms are trumps and should always be prioritised over other harms. Applying these arguments to the case of PAS, I am able to defend a claim in chapter six that the permitting of PAS is not a desirable extension of patient choice. I argue both that the types of weak character I describe and being adversely influenced by the context of choice are relevant in a situation where PAS is made permissible. Consequently, should PAS be made permissible, certain patients may choose PAS and be harmed by wrongful death. However, should PAS not be made permissible, then selected patients who are suffering unbearably at the end of their lives, and whose suffering is unrelievable, may be harmed because they do not have the option of ending their suffering by dying. So, in order to decide if PAS should be made permissible, the harms that will accrue to certain patients, should it not be made permissible, have to be weighed against the harms that will accrue to certain patients, should it be made permissible. Drawing on the arguments in chapter five, I argue that the harm to patients who suffer wrongful death as a result of choosing PAS is a trumping harm. Furthermore, despite its gravity, unbearable suffering is relevantly distinct from wrongful death. This is because unbearable suffering in the context of having a live option of PAS does not have qualities, namely of foreclosing autonomy or rendering patients unable to meet the psychological requisites for survival, that would put it on a par with wrongful death. So, the harm to patients who suffer wrongful death as a result of choosing PAS, should it be made permissible, trumps the harm to patients who are suffering unbearably at the end of life and who are denied PAS, should it not be made permissible. Based on all these arguments, I conclude that the permitting of PAS is not a desirable extension of patient choice.

Chapter 1. The Harm of Additional Options

1.1 Abstract

This chapter argues that certain types of weak character and what I term the "context of choice" may give rise to two harms to an agent who is presented with an additional option. The first of these two harms is accrued by an agent who is given the opportunity to exercise her weak character, and the second harm occurs because the context of choice affects the way the choice situation appears to an agent, with the result that the salience to her of certain aspects of the choice situation is harmfully altered. Dworkin, G. (1988) and Velleman (2015) have also argued that an agent may potentially be harmed in several different ways if she is offered an additional option. I analyse Dworkin's and Velleman's nine harms and defend a claim that each of them is different from the two additional harms that I identify. They are different because it is not a necessary feature of Dworkin's and Velleman's harms, either that the agent has a weak character of the types I defend, or that the context of choice is adversely configured. However, both of the harms I defend may additionally occur in the types of case described by Dworkin and Velleman.

1.2 Introduction

The line of argument that an agent may potentially be harmed by being offered an additional option is relevant to my thesis because, should PAS be made legal, then certain patients would be offered PAS as an additional option—an option that is not currently available in the UK. If this is a case in which having an additional option may be worse than not having an additional option, a case where having the additional option is harmful, at least to some people, then this harm should be considered when deciding if options should be extended in this way[3].

A first step in the argument in defence of the existence of harms arising from additional options is to explain the nine ways in which Dworkin, G. (1988) and Velleman (2015) claim that having fewer options can be better than having additional options. This

[3] I also acknowledge in chapter six that should PAS *not* be made legal then this may potentially harm an agent at the end of her life who is suffering unbearably and whose suffering is also unrelievable. An agent in this situation may potentially be harmed because her unbearable suffering cannot be ended as a result of her death by PAS.

analysis will illuminate the types of case in which an agent may be potentially be harmed by being offered an additional option, and also indicates that there are two further aspects of choice situations that may result in harm accruing to an agent. The first of these aspects is relevant features of an agent's character and the second is adverse features of what I term "the context of choice". Neither Dworkin nor Velleman include accounts of relevant features of the agent's character or the context of choice in their arguments.

The relevant features of an agent's character and the context of choice are applicable here because two significant components of any choice situation will be, first, the character of the agent who is choosing, and second, the context in which she is choosing. A full development of my thesis will require a detailed consideration of each of these components, but for the present I claim that with respect to the character of the agent it is possible to identify relevant features that are types of weak character. I also note that what I call the context of choice consists in selected features of a choice situation that are external to the agent.

So, to develop my position I will need to give preliminary accounts of the relevant features of an agent's character and of the context of choice that make it possible for harm to result if an additional option is presented to an agent. My account of character and selected types of weak character will be fully defended in chapters two and three, and my account of the context of choice will be fully defended in chapter four, but for the purpose of this chapter I will rely on working accounts.

I will argue that harm to an agent who is presented with an additional option is dependent, at least in part, on her having certain relevant features of character that make her susceptible to harm. An agent is more susceptible to harm from having additional options if her character is not robust—that is, first, if any part of her capacity for judgement and choice is weak[4], second, if she has an unduly low self-esteem[5], and third, if she has undue lack of confidence in her judgements. So, to illustrate the last of these, one way in which an agent may be harmed if she is presented with an additional option is by experiencing additional difficulty in taking responsibility for her choice as a result of undue lack of confidence in her judgements based on her values.

---

[4] In this chapter I use lack of robustness of character and weak character interchangeably.
[5] Having low self-esteem could also be expressed as being unduly self-deprecating.

I will then argue that, at least pre-theoretically, the features of the context of choice that make it more likely for harm to result, if an additional option is presented, are relevant features of the choice situation that are external to the agent and which may adversely influence her decision-making process. Features of a choice situation external to the agent that may affect her decision-making process are those that are salient to her; they may include qualities of the options themselves, such as their value to her, their complexity, and setting. The setting may consist in selected aspects of the choice situation including features of other people and the physical environment. For example, the behaviour of other people in the choice situation may be relevant, as it may reveal their own valuations of the available options. All these factors may adversely influence the agent in a choice situation. I note, too, that there is an important normative aspect to the context of choice that I develop fully in chapter four. In brief for now, the normative aspect of the context of choice consists of features of the choice situation that are external to the agent and that either should or should not be salient to her.

Having given preliminary accounts of types of weak character and the context of choice, I next defend a claim that each of these notions can explain harms to an agent in a situation where she is presented with an additional option. Furthermore, the two harms that I identify are distinct from Dworkin's and Velleman's harms. The structure of the argument in defence of my two additional harms is to revert to the harms described by Dworkin and Velleman. I argue, first, that Dworkin's and Velleman's harms are not dependent on the agent having a weak character of the types I describe or on adverse features of the context of choice. Second, I will argue that the additional harms that I identify may additionally occur in the cases used by Dworkin and Velleman to explain their nine harms.

1.3 Dworkin's and Velleman's Harms from Additional Options

Between them, Dworkin, G. (1988) and Velleman (2015) identify nine distinct possible ways in which harm may accrue to an agent who is offered an additional option. The harm that results from the offer of a new option is normally experienced by the agent, and this is the main focus of my dissertation, but I note that society in general may also be harmed by giving agents a new option[6]. My purpose in this section is to outline

[6] One way in which society as a whole may be harmed in situations where there are additional options is as a result of members of a society having less things in common with each other and which they

Dworkin's and Velleman's harms before returning to them in a later section where I defend a claim that my two additional harms are distinct from theirs.

1.3.1 Dworkin's First Harm: "Decision-making Costs"

Dworkin identifies seven harms from additional options and the first of these is the harm that results from "decision-making costs" (D1) (Dworkin, G., 1988, p.66). In this type of case, harm results to an agent from extra costs in terms of the intellectual effort involved in assessing, weighing and deciding between the available options. I note here that an agent may accrue decision-making costs if there is one additional option or if there are multiple new options added to the choice set. An example of one additional option which may incur extra decision-making costs for a patient is a complex new surgical procedure that is relatively untried. Alternatively, a choice set may be complex because it consists in a large number of less intrinsically complex and dissimilar items, such as Scitovsky's "ten different kinds [...] of shirt" (Scitovsky, cited in Dworkin, G., 1988, p.72)[7]. Choice sets with a greater quantity of dissimilar items may present problems to an agent in the same way as smaller sets consisting in intrinsically complex items may do. Adding ninety shirts to Scitovsky's ten shirts is an example of a multiplicity of new options that may incur extra decision-making costs.

I will argue in a later section that the harm to an agent in "decision-making costs" occurs regardless of any facts about her character or the context of choice. Before I do this I will outline a preliminary account of certain types of weak character and a pre-theoretical account of the context of choice. For the moment, however, I will progress with brief accounts of Dworkin's remaining six harms and Velleman's two harms.

1.3.2 Dworkin's Second Harm: "Responsibility for Choice"

The second way, described by Dworkin, in which an agent may be harmed if she is presented with an additional option is termed "responsibility for choice" (D2) (Dworkin, G., 1988, p.67). The harm in D2 is from an agent being held responsible for her decision by other people (either in society or by the law). Ronald Dworkin distinguishes two senses of responsibility: first, a moral sense involving responsible or irresponsible

---

can share. So, for example, if there are fewer TV programmes to watch, then it is more likely that people will be able to share the experience of watching these programs and mutually benefit from the process of sharing their experiences. I am grateful to Professor R. Chadwick for giving me this example during a research seminar.

[7] I have been unable to find Scitovsky's quote in the source cited by Dworkin.

behaviour, and second, a relational or causal sense in which "someone is or is not responsible for some event or consequence." Both senses of the word responsible can apply in discussions about responsibility for a decision. Once an agent has made a decision in any particular choice situation she is "responsible [in the second, relational, sense] and [can] be held responsible for events that prior to the possibility of choosing were not attributable to [her]" (Dworkin, R., 2011 , p.102). Being responsible in the relational sense of the word creates "the pressure (social and legal) to make 'responsible' choices" in the first, moral, sense (Dworkin, G., 1988, p.67). How can responsibility for a decision be harmful?

Dworkin gives as an example of D2 a case in which a woman is given the additional option of selectively aborting her foetus if it has Down's syndrome. If a woman is presented with the option of aborting her foetus then she is both causally and morally responsible for her decision in the eyes of society. Responsibility for the decision she makes in the *moral* sense, whether her decision is to have an abortion or to continue with the pregnancy, is rightly considered to be a very weighty matter. So the harm to the agent results from being viewed by other people as having the moral responsibility to make a grave decision correctly.

1.3.3 Dworkin's Third Harm: "Pressure to Conform"

"Pressure to conform" (D3) (Dworkin, G., 1988, p.68), is harm to an agent who is subjected to pressure to choose a new option. Pressure to conform may be harmful to her because she does not want to choose the new option. Dworkin describes two types of pressure to conform: the first of these is indirect pressure to choose a new option as a result of this choice being a social norm and the second, very briefly, is harm to an agent who is subjected to direct pressure to choose a new option. For the purposes of my overall argument, however, I am not interested in situations where people are put under direct pressure, perhaps by being threatened. More relevant for my arguments is the social norm variant of D3, in which a new option is made widely available and a significant number of people choose it. In this type of case an agent who is offered the new option does not value it, but is aware that other people do value it and as a result have chosen it. The knowledge that other people have chosen the option puts the agent under pressure to make the same choice, so that her choice can be in line with the choices of other people.

Two of Dworkin's examples of pressure to conform (D3) are pressure on an agent to move into a co-ed dormitory and pressure to selectively abort foetuses of a certain gender when foetuses of that gender are less valued by society. Dworkin (1988) also quotes Schelling (1960), who gives a further example of pressure on an agent to settle a feud by duelling in a culture where duelling is the normal way of resolving such disputes[8]. In D3, there need be no intention on the part of the other norm-making people to pressurise the agent; people who move into the co-ed dorm, abort or duel do not necessarily intend that other agents follow suit.

One significant aspect of D3 is that people in the choice situation other than the agent behave in certain relevant ways. These people are part of the context in which the agent is choosing, so they are part of the context of choice. However, in my account of the context of choice, we shall see that its harmful effects are accrued by the agent as a result of various alterations in salience to the agent of such features of the choice situation.

1.3.4 Dworkin's Fourth Harm: "Exercise of Choice"

Dworkin calls his fourth harm from additional choice "exercise of choice" (D4) (Dworkin, G., 1988, p.69) because it describes an agent whose choice (her "exercise of choice") is altered in a harmful way. An agent in D4 undergoes a change in motivation, though this change in motivation is not as a result of external pressure of any type, as in D3. In D4 the agent is harmed because she no longer chooses a pre-existing option that she valued once a new option has been offered alongside it: her motivation to choose the pre-existing option is extinguished. This negative effect on the agent's motivation towards the pre-existing action results from an unappealing change (to her) in the pre-existing option when the new option is offered. An important aspect of D4 is that the new option has extensive features in common with the pre-existing option, and this partially explains why the agent's motivation towards the pre-existing option changes.

Dworkin gives the case of altruistic blood donation, as described by Titmuss (1972, cited in Dworkin, 1988, p.70) as his main example of D4. Titmuss claims that when a market for blood giving is created, thereby making available the additional option of

---

[8] A further example of D3 could include pressure on a patient to request PAS if it were legalised and became a social norm, especially within a population of people with the same disease, e.g. motor neurone disease. As I have said, I save my discussion of PAS for chapter six, after I have laid down the necessary supporting arguments.

selling your blood, it becomes more difficult for an agent to give her blood altruistically. It can be seen that in the new scheme, where selling blood becomes an option, there is no loss of freedom to give altruistically. However, in order to give blood altruistically after the introduction of the new option the agent has to refuse payment for her blood. The addition of blood selling to the existing option alters the nature of altruistic giving and thereby causes a negative change in the agent's motivation towards the latter, because she *feels* altruism is no longer necessary. Singer goes further, and says:

> "a commercial system may discourage voluntary donors. It appears to discourage them, not because those who would otherwise have made voluntary donations choose to sell their blood instead if this alternative is available to them (donors and sellers are, in the main, different sections of the population), but because the fact that blood is available as a commodity, to be bought and sold, affects the nature of the gift that is made when blood is donated.
>
> [E]ven if these people had the formal right to give to a voluntary program [sic] that existed alongside commercial blood banks, their gift would have lost most of its significance.... The fact that blood is a commodity, that if no one gives it, it can still be bought, makes altruism unnecessary[.]" (Singer, 1977, cited in Dworkin, 1988, p.71.)

Singer's claim is that altruism itself becomes unnecessary after the introduction of the new scheme. However, an altruistic act does not depend for its possibility on whether or not the act may also be associated with a payment. The possibility of altruism merely depends on the agent being able to give selflessly.

1.3.5 Dworkin's Fifth Harm: "Increased Choices and Welfare Decline"

Dworkin's fifth harm from additional choice is "increased choices and welfare decline" (D5) (Dworkin, G., 1988, p.73). In D5 an agent is placed in a worse position if they have the choice of a new option than they would be if they did not have this choice. The agent does not need to choose the new option to be harmed, but is exposed to a risk of future harms resulting from the actions of other people if she is in a position where the new option is available to her. Furthermore, the agent in D5 values the new option despite there being a possibility of loss of welfare to her as a result of the threat posed by the other people's actions. In this case, then, the actions of these other people are linked with the agent's choice, and have a harmful effect on her.

Three examples of the linkage in D5 between the agent and the other people involved in the choice situation are given by Dworkin. These are, first, a "bank teller [who] knows the combination to the safe, [and] can be threatened into opening it"; second, a "doomsday machine that responds automatically to [an attack]" thereby making threats

of retaliation "more credible"; and third, prisoners' dilemma situations where the best outcome for both parties is obtained if they are prevented from enacting their preferred choices (Dworkin, G., 1988, p.74). In each of these cases the agent prefers to have the choice of the additional option, but having this option exposes her to harm from the actions of other people.

1.3.6 Dworkin's Sixth Harm: "Morality and Choice"

Dworkin's sixth potential harm from additional choice appears under the heading "morality and choice" (D6) (Dworkin, G., 1988, p.75). The potential harm in D6 results from creating a new option, that is available to some people only. One of Dworkin's examples of this type of harm is having the exceptional option of being able to buy one's way out of conscription, in a society where conscription is valued. In D6, an agent to whom the new option is offered receives it on an exceptional basis and so, Dworkin suggests, should voluntarily restrict her access to the option in order to avoid the unfairness in having it. If the agent were to take up the new option, then society would be harmed as a result of loss of community solidarity. The agent herself may gain as a result of having her preferred option, but she could also be harmed as a result of potential alienation from societal values. So if the agent takes up the new option then she may suffer a form of welfare decline, as in D5.

Dworkin, however, claims that the harm that results from having the additional option in D6 occurs whether or not it is taken up by the agent: "it is already morally significant that he has the choice—whether or not he intends to make use of it. It is his having the choice, whereas others do not, which is ruled out on moral grounds" (Dworkin, G., 1988, p.75). So there is no specific harm to the agent in D6, but society is harmed as a result of loss of community solidarity.

1.3.7 Dworkin's Seventh Harm: "Paternalism and Choice"

"Paternalism and choice" (D7) (Dworkin, G., 1988, p.76), is Dworkin's seventh and final type of harm arising from the offer of an additional option. The harm in D7 is accrued by an agent who recognises in advance that she may be motivated to choose a harmful new option. In this case, an agent who wishes to avoid being placed in a position where she may undergo a change in motivation towards a harmful new option agrees in advance that her access to it should be restricted. Her access could be limited in a number of different ways: for example, an agent may ask her friends to keep an eye

on how much she is drinking at a party and limit her access to alcohol if she is showing signs of becoming intoxicated, the state may contract with its population to set limits on specific behaviours such as speeding or drug-taking, or access to high places from which the suicidal may throw themselves may be limited. In other words, an agent agrees in advance to a restriction in the range of options available to her so that she is not tempted to fall into behaviour she wishes to avoid.

D7 appears to trade on the agent being aware of a potential weakness in her character, namely one that may result in her choosing a harmful option. However, I will argue below that the types of weak character I identify are either broader than the type in this case or different from it. Having explained Dworkin's seven harms that arise when an agent is offered an additional option, I now explain Velleman's two harms from an additional option.

1.3.8 Velleman's First Harm: "Preference for the Default"

In the context of an essay on PAS, Velleman describes two harms that can accrue to an agent who is offered an additional option. The first of these, V1, is the harm arising from "preference for the default" (Velleman, 2015, p.10). The harm to an agent in V1 occurs because she prefers a pre-existing (default) option. However, she has to choose and justify choosing this pre-existing option when an additional option is offered to her. Of course she had no need to justify choosing the pre-existing default option by comparison with the new option *prior* to the new option being made available. In V1, the agent and the people making the offer differ over their valuation of the new option: the agent does not value the new option before it is offered, but, since it would be unusual for anyone to offer an option that they do not value, we can assume that the people making the offer do value it.

One example of V1 given by Velleman is of an agent who is invited to a dinner party, but who would prefer to stay at home instead[9]. Before the option of attending the dinner party was offered, the agent had no need to justify her preference for staying at home, but once the invitation is made, the agent then has to justify her choice. So the harm to the agent in V1 is that she must now do the chore of refusing the invitation.

---

[9] Velleman's other example of V1 is of an offer of PAS. I apply the arguments in chapters one to five to PAS in chapter six, so I save my discussion of PAS until then.

Furthermore, in refusing the invitation, the agent has to reveal her preference to the person who gave the invitation.

1.3.9 Velleman's Second Harm: "Harmful Implications"

Velleman's second harm from additional choice may be accrued by an agent who is offered an option that may bring with it harmful implications for her (Velleman, 2015, p.11): call it "harmful implications" (V2). In V2, as in V1, the new option being offered is valued by the person making the offer—he feels it will be valuable to the agent. Also as in V1, the new option is not valued by the person to whom it is offered—she does not feel that she needs it.

Velleman's example of V2 is of a student who is offered remedial instruction by her lecturer. The student may be harmed by the offer of assistance because she does not, in fact, need assistance, but wonders in the light of the offer if she may do so. In this example, the source of the offer of assistance is likely to be authoritative and the harmful implication of the offer is that the student is performing poorly in her studies.

Table 1: Dworkin’s and Velleman's Harms from Additional Choice with Examples

		Term.	Example.
D1.	Dworkin.	Decision-making costs.	Choosing shirts, pension schemes.
D2.		Responsibility for choice.	Selective abortion of a foetus with Down’s syndrome.
D3.		Pressure to conform.	Peers of the agent choosing to stay in a co-educational dorm.
D4.		Exercise of choice.	Effects on altruistic blood donation if a scheme for selling blood is made available.
D5.		Welfare decline.	Bank teller who knows the combination to the safe.
D6.		Morality and choice.	Option for some people of buying one’s way out of conscription in a context where conscription is valued by the majority.
D7.		Paternalism and choice.	Options of harmful drugs, driving without seat belts etc.
V1.	Velleman.	Preference for the default.	Invitation to dinner.
V2.		Harmful implications.	Offer of remedial instruction by a teacher to a student.

1.4 Summary for Dworkin and Velleman.

In this section I have explained Dworkin’s and Velleman’s nine harms accrued by an agent who is offered an additional option. In my discussion of two of these harms I briefly mentioned the character of the agent (in paternalism and choice, D7) and the context in which the agent is choosing (in pressure to conform, D3). D7 and D3 thereby give us reason to think that there may be two aspects of a choice situation, namely an agent’s character and the context of choice, that are relevant in this type of situation. Furthermore, the notions of character and context may also explain further harms to an agent who is presented with an additional option that are not identified by either Dworkin or Velleman. In the next section, then, I will defend a claim that there are two harms that may be accrued by an agent if she is presented with an additional option which are distinct from Dworkin’s and Velleman’s harms. First, an agent may be harmed in this type of situation if she has certain types of weak character, and second, she may be harmed if the context of choice is adversely configured.

1.5 A Preliminary Account of Relevant Features of an Agent's Character

I now turn to the first of the two key aspects of choice situations which I claim may influence an agent to be harmed when she is presented with an additional option. As I have indicated, the first of these is relevant features of her character. I claim that an agent is more likely to be harmed when she is presented with an additional option if her character lacks robustness or is weak in some way. The offer of the additional option presents her with an opportunity to *exercise* her weak character. Weakness may be manifested in the parts of her character concerned directly with decision-making, as well as in the parts of her character indirectly related to decision-making, such as those that influence her ability to take responsibility for her choices.

First, a lack of robustness in decision-making implies that the agent has a reduced capacity for judgement and choice. I will elaborate on the capacity for judgement and choice in chapters two and three[10], but for the moment I will use an outline account from Buchanan and Brock (1990) because this covers the necessary ground for my immediate purposes. Buchanan and Brock claim that the capacity for judgement and choice is a process that can be divided into two sub-capacities: "[first] the capacity for understanding and communication and [second,] the capacity for reasoning and deliberation" (Buchanan and Brock, 1990, p.23). An agent who lacks robustness in judgement and choice may therefore have problems with any or all of the understanding, communication, reasoning and deliberation which are required for making choices. Most people have a degree of deficiency in these features of decision-making, so weak character will reflect the degree of this deficiency and its relevance for the agent in choice situations. The relevance of weak character will also depend in part on the particular decision being made.

In a situation where an agent is presented with a range of options to choose from, the process of understanding the options depends first on being able to acquire information and then on having the necessary cognitive skills to comprehend and assimilate the information acquired. Information about the options on offer may be acquired by being directly perceived by or being communicated to the agent. So an agent who has

---

[10] My accounts of character and types of weak character in chapters two and three will be in the Aristotelian tradition.

difficulties with perception and communication in any given choice situation probably will not have robust decision-making skills. If the agent is able to acquire and assimilate the necessary information about the options on offer, she then has to reason about them and to deliberate, in order to come to a preferred option.

Difficulties in the process of reasoning and deliberation, which I have run together for the purposes of this chapter, may reflect aspects of an agent's character in the following three ways. First, the agent may have difficulty weighing different options on offer against one another so that she can come to a preferred option. Second, if the agent is not accurate in her assessment of her needs or abilities when faced with a choice situation, then she will not be certain which of the different options on offer are of most value to her, and which can thereby best suit her needs and abilities. Some of the options on offer may suit an agent with poor abilities or extensive needs, and an agent who does not have extensive needs but who overestimates them may wrongly believe that these are the options that she should give preference to. Third, an agent may have an undue lack of self-confidence in her judgements, and this may cause her to feel uncomfortable with responsibility for the decision she has made.

In this preliminary account, I have identified three different features of an agent's character that are relevant in choice situations, and which I defend in chapter three. These are types of weak character: first, difficulties with weighing and choosing, second, not being able to make an accurate assessment of one's own needs and abilities, and third, having an undue lack of confidence in one's judgements. Importantly, in situations where an agent with a weak character is offered an additional option, the offer gives her the opportunity to exercise her weak character.

Having given an outline of the relevant features of an agent's character which make her susceptible to harm in a situation where an additional option offered, I now describe configurations of the context of choice which may result in her being harmed. I will fully defend my account of the context of choice in chapter four and so give at this point only a pre-theoretical account sufficient for my immediate purposes. The context of choice includes the features of any choice situation external to the agent which may influence her decision-making.

1.6 A Pre-theoretical Account of the Context of Choice

Features of the choice situation that are external to an agent and which may have an influence on her decision-making constitute the context of choice. These external features may consist in, first, relevant aspects of the other people in the choice situation, and second, relevant aspects of the physical environment. The relevant aspects of other people in the choice situation can be subdivided into the values of the other people, their intentions and the effects of their actions. The relevant aspects of the physical environment can be divided into qualities of the options, such as their value or lack of value to the agent, and the complexity of the options. I will further develop my account of the context of choice by arguing in chapter four that it consists in features of a choice situation that are either salient to an agent, or should or should not be salient to her.

People other than an agent may be involved in a choice situation in one of three ways: first, they may be responding to a situation where a new offer is made to an agent but not offered to them—for example, they are in a shop with an agent who is deciding how to spend some lottery winnings—or second, they may have been offered the new option themselves—for example, they are adjacent to an agent who is deciding which sweets to buy at a supermarket check-out—or third, they themselves may be offering the new option, for example, they are serving behind the till at the shop. In each of these cases, the intentions and actions of other people in the choice situation can have an effect on the agent who is faced with a new option, and she may potentially be harmed as a result.

First, other people who are responding to a choice situation where a new offer has been made to an agent, but who are not being offered the new option themselves, may have a specific interest in the outcome of the agent's decision making process and so may intend to pressurise the agent's decision-making process in one way or another. Alternatively, they may respond in a way that creates pressure on the agent without intending to do so: their behaviour may influence the agent to make a certain decision, perhaps by leading her to believe that she should choose a certain option which she had no prior reason to choose.

Second, other people may be involved in a choice situation because they too have been offered the new option. If sufficient numbers of other people respond to the offer of a new option in the same way as each other, then this may create a social norm—a state of affairs that can influence the behaviour of an agent in the same circumstances if she

is aware of this state of affairs. Consider the supermarket example (above), where the agent may be influenced by a "consumption norm" (Elster, 1989, p.100)—for example the tendency of the majority of people at large (i.e. not just in the shop with the agent) to buy a new brand of chocolate. The agent may also be influenced by smaller numbers of people, or exemplars.

Thirdly, other people involved in a choice situation can have an effect on an agent through the way in which the option is offered. The offer of a new option to an agent may itself influence her to choose it. One possible way in which she may be influenced is if it is clear to her that the people offering the option value it highly. In this case, the agent may be swayed towards choosing the new option. As an example here, consider the cashier in the shop (above) who describes a new chocolate bar to the agent in glowing terms thereby causing her to be swayed towards buying it. Conversely, the offer of a new option that is not valued by the person offering it may have the effect of deterring the agent from choosing it.

The value to the agent of the new option also has an effect on the choice she makes. She is more likely than not to reason that she should select a new option that she values. However, the agent's values may conflict with the values of other people involved in the choice situation. This could result in harm to her if she is not confident in judgements that flow from her values in the face of contrary values held by other people. It should also be noted that the agent's judgements reflect her character, so there is an intersection of this feature with features of the agent.

Next in my preliminary account of the context of choice in which harm may result to an agent is the complexity of the options on offer. A single additional option in a choice set may be highly complex in itself. It may have features that make it difficult for the agent to acquire and assimilate relevant information about it, and then to reason and deliberate before making and communicating her choice. Alternatively, a choice set may be complex because it consists in a large number of less intrinsically complex and dissimilar items such as Scitovsky's "ten different kinds [...] of shirt" mentioned above. (Scitovsky, cited in Dworkin, G., 1988, p.72). Choice sets with a greater quantity of dissimilar items will present problems to some agents in the same way as smaller sets consisting in intrinsically complex items.

A final observation about the relevant features of both the context of choice and the agent's character, that make harm more likely to result in a situation of additional choice, is that these features may be linked. So, if an agent in a choice situation where additional options are on offer does not have a robust character then she may be susceptible to various harms, and these harms will be more likely if there are also adverse features of the context of choice. It is also possible, then, that an agent with a more robust character may be harmed if she is choosing in a choice situation where the context of choice is especially adverse. Conversely, even if the context of choice is optimal, an agent with lack of robustness of character may still be harmed. For example, the behaviours of people other than the agent who are offered an additional option—part of the context of choice—may influence the behaviour of the agent who has a relevantly weak character. And even if the agent's character is robust, certain behaviours of other people in the choice situation may have an adverse influence on her. Finally, even if the behaviours of other people in the choice situation are optimal in respect of choice-making, an agent who lacks robustness of character may still be harmed.

Having given this preliminary account of the two key features of choice situations that I claim may result in harm to an agent who is offered an additional option, I now defend two further claims. First, weak character and the context of choice can result in harms to an agent who is presented with an additional option that are not identified by Dworkin or Velleman. In order to defend this claim I revisit Dworkin's and Velleman's harms and show that they do not depend on either weak character of the types I describe or on adverse features of the context of choice. Second, I argue that the two new harms that I identify may arise in some of the cases used by Dworkin and Velleman to illustrate their harms.

Table 2: Dworkin's and Velleman's Harms from Additional Choice

		Term for Harm
D1.	Dworkin.	Decision-making costs.
D2.		Responsibility for choice.
D3.		Pressure to conform.
D4.		Exercise of choice.
D5.		Welfare decline.
D6.		Morality and choice.
D7.		Paternalism and choice.
V1.	Velleman.	Preference for the default.
V2.		Harmful implications.

1.7 Weak Character and Context of Choice: Dworkin's and Velleman's Harms Revisited

1.7.1 "Decision-making costs"

Dworkin terms his first harm from additional choice "decision-making costs", D1 (Dworkin, G., 1988, p.66). One case in which this type of harm is relevant is where an agent is choosing a suitable pension scheme, e.g. Cronqvist and Thaler (2004) and Iyengar (2011). There may be considerable decision-making costs involved in assessing and weighing all the different pension schemes on offer and, importantly, these costs will be accrued by the agent even if she has robust decision-making capabilities. Consider a case in which a pension provider offers its members a new collection of pensions to choose from. In this case, too, an agent with robust decision-making skills will face the potentially harmful burden of assessing the new schemes. Furthermore, in a case where an agent is offered new pension schemes there need not be any harmfully distorting effect of this offer on the context of choice, the second aspect of the choice situation that I claim may result in harm to an agent facing additional options. The offer of the new schemes need not result in some aspects of the choice situation being salient when they should not be salient or *vice versa*. So Dworkin's first harm is not reliant on either weak character or on adverse features of the context of choice.

If, now, we consider the role of weak character in the pensions case, it can be seen that this feature may result in an additional harm to the harm described by Dworkin. In

virtue of a new offer, such as a pension scheme, an agent may be given an opportunity to exercise her weak character. She may, first, exercise her weak decision-making skills and be additionally harmed by the consequent exertions. Second, she may select a scheme that does not match her needs as a result of unduly low self-esteem. And third, she may struggle with the process of choosing as a result of undue lack of confidence in her judgements arising from her values. So the offer of the new scheme may harm her in any of these three ways as a result of her exercising her weak character.

The harm that may result from adverse features of the context of choice may also occur in cases used to illustrate D1. By way of a reminder, adverse features of the context of choice (which I fully defend in chapter four) result in the salience to an agent of features of the choice situation being harmfully altered. The result of this is that the agent either finds aspects of the choice situation salient that she should not find salient, or does not find salient features of the choice situation that she should find salient. An example of this type of harm is the agent choosing a pension scheme as a result of finding it salient when she should not have found it salient. This harmful change in salience to the agent may occur as a result of the actions of other relevant people in the choice situation. If, for example, a new pension scheme is very popular, then this may render the scheme more salient to the agent when it should not be more salient to her.

So far in this section I have revisited the first of Dworkin's harms in order to defend a claim that it does not depend either on an agent having a weak character or on adverse features of the context of choice. I also argued that the harms of weak character and context of choice may additionally occur in cases used by Dworkin to illustrate his first harm. So the harms resulting from weak character and adverse features of the context of choice are distinct from Dworkin's first harm. In the remainder of this section I follow the same structure to argue that none of Dworkin's remaining harms or Velleman's harms depend either on my weak character types or the context of choice. Furthermore, the harms of weak character and context of choice may additionally occur in cases used by Dworkin and Velleman to illuminate these remaining harms. So I will conclude that the harms of weak character and context of choice are distinct from all of Dworkin's and Velleman's nine harms.

1.7.2 "Responsibility for Choice"

The harm to an agent in "responsibility for choice", D2 (Dworkin, G., 1988, p.67), occurs in situations where she is harmed by having to take responsibility for her

decision. As I have said, the case used by Dworkin to illustrate this harm is of an agent who is held responsible in society for the grave decision whether or not to have an abortion if she conceives a child with Down's syndrome. Furthermore, this is a choice situation where there are conflicting views at large, and this is likely to add to the harm experienced by the agent as a result of her having to take responsibility for her decision. In this case it can be seen that, first, an agent with a robust character may still potentially be harmed as a result of having to take responsibility for her decision. Clearly in other cases this harm will depend in part on the seriousness of the decision being made, but the point is that harm may result to the agent if the decision is a sufficiently grave one, and even if her character is robust. It could be objected here that an agent with a robust character could just "shrug off" a grave decision. However, this would not be a normal reaction to a weighty decision of the type I have described since it would suggest that the agent is not taking the decision seriously. Second, in this example of a choice situation, it is clear that it is not necessary in order for harm to result to the agent that there is any adverse influence of the context of choice.

Harms that may be accrued by an agent as a result of adverse effects of the context of choice result from changes in salience to the agent of features of the choice situation. In D2, on the other hand, in order for the agent to be harmed she must merely be aware of other views at large. None of these other views need be either salient to the agent when they should not be, or not salient to her when they should be, in order for harm to result. If there *is* an adverse effect of the context of choice, however, and the agent for example finds certain views of other people more salient than she should do, then this will also result in harm to her. This harm, since it results from adverse features of the context of choice, is distinct from Dworkin's harm. Additionally, if the agent has a weak character of the types I describe then this may also result in harm to her in cases that Dworkin utilises to illustrate D2.

In the types of cases that Dworkin utilises, an agent may additionally be harmed if she has particular difficulty taking responsibility for her decision. She may have particular difficulty taking responsibility for her decision if she has an undue lack of confidence in the judgements that flow from her valuation of the option—the third type of weak character. If she is not good at taking responsibility for her choice because she lacks confidence in the judgements that flow from her values, and especially if her values conflict with those of the majority, then she may doubt that her choice is the correct one, and may be harmed as a result.

Other people whose values are in conflict with the agent's values could be thought of as unintentionally applying pressure to the agent—pressure to choose in line with the other people's values. As we have already seen, pressure on the agent is Dworkin's third harm: "pressure to conform", D3 (Dworkin, G., 1988, p,68), to which we now return.

1.7.3 "Pressure to Conform"

An agent who is faced with a range of options may be harmed by feeling pressure to make certain choices if there is a disparity between her values and the values of other people in the choice situation. As with the harm of "responsibility for choice", the harm of D3 may be greater if the decision being made is graver. Furthermore, even an agent with a strong character will be aware of this pressure and may potentially be harmed by it. As we saw, one of Dworkin's examples of D3 is harm due to pressure to use a co-ed dorm. In this case, an agent whose values dictate that she should not use a co-ed dorm, and whose character is robust, may be harmed by pressure from the many people who have chosen to use a co-ed dorm. Furthermore, the agent in this case may experience pressure on repeated occasions, for example each time she is asked where she is living if the person who asks this knows about the different options on offer. So having a weak character is not a requirement for the harm in D3. Similar to the harm in D2, the context of choice also does not have to be adversely configured for the agent to be harmed in D3: there is no harmful alteration of salience of aspects of the choice situation since the agent is merely aware of other people's values. If, however, the values of other people in the choice situation are salient to the agent when they should not be, then a different type of harm arises in this case—one due to adverse features of the context of choice.

I have explained that the harm in D3 arises if an agent's values are in conflict with the values of a sufficient number of other people whose values are known to the agent. If an agent in this situation has a weak character then the harm from pressure to conform will be increased. This is because her weak character may make her more susceptible to indirect pressure from a social norm. The type of weak character that is relevant here is the agent having an undue lack of confidence that her values are well founded. It should be noted that the harm when "weak character" is applied to a case used by Dworkin to illustrate D3 intersects with the harm of D3 itself . Both types of agent, i.e. with robust or weak characters, experience the same pressure but to different degrees. Having a weak character is not a necessary feature of D3, however the types of choice situation

relevant to D3 present to an agent with a weak character the opportunity to exercise this weakness.

1.7.4 "Exercise of Choice"

The harm to an agent in "exercise of choice", D4, occurs when, following the introduction of a similar option, the agent is no longer motivated to choose the option that she previously valued. I argued above, against Singer (1977), that, in the specific case of creating a market for blood giving, the possibility of altruistic donation persists. So when a market is created for blood giving this does not mean that the agent no longer has the option of altruistic donation; she merely ceases to feel that altruistic donation is as necessary as it was before. Her assessment of the necessity of altruistic donation in this case is not reliant on her having a weak character. Nor is the harm in cases of D4 such as this one reliant on an adverse effect of the context of choice. More specifically, the offer of the new option does not have to harmfully alter the salience to the agent of any of the features of the context of choice in order for her to accrue harm.

However, should the agent in this case have a weak character of the types I have described, then this may result in harm to her. A weak capacity for judgement and choice may result in the agent struggling with her decision about whether to continue giving blood altruistically. Being unduly self-deprecating may result in her being less likely to stop giving blood altruistically, since altruism itself is reliant on a certain selflessness. Finally, if the agent has an undue lack of confidence in her judgements based on her values, then she may be more ready as a result of the new option to exercise this weakness and to choose not to give altruistically.

My second harm, from adverse features of the context of choice, may also apply in the blood donation case. As I have said, in order to have an adverse effect, the context of choice must result in the agent finding aspects of the choice situation salient that she should not find salient and *vice versa*. So, for example, the agent may find the new option (paid blood donation) more salient than she should do, and this may result in her being less likely to select the pre-existing option (altruistic donation). This distorting effect of the context of choice was not necessary for the harm to result in D4, but when there is a distorting effect of the context of choice then this results in a similar type of harm to the agent. In other words, there is an intersection of these two harms.

1.7.5 "Increased Choices and Welfare Decline"

In Dworkin's fifth harm, "increased choices and welfare decline", D5 (Dworkin, G., 1988, p.73), the agent has an additional option that exposes her to potential harm from another agent or agents. As we saw, Dworkin describes a bank teller whose knowledge of the combination to the safe exposes her to the potential harm of being pressurised into opening the safe. The harm to the bank teller in Dworkin's example would be present even if she had a robust character. The bank teller knowing the combination to the safe does not necessarily reflect her having a weak character, as knowing the combination is a normal part of her role. If the bank teller's character is weak, however, then the availability of certain other options may result in her exercising her weak character. She may, for example, choose an option that presents itself, such as carrying on her person the key to the back door of the bank to enable her to take opportunistic breaks. In this second case she is not only exposed to potential harm as a result of knowing the combination to the safe, but also from possessing the means to a quick escape route for the robbers. This is because of her lack of awareness of the potential harms inherent in her choice. So, a type of weak character relating to capacity for judgement and choice could lead an agent to choose a harmful new option.

There is no adverse configuration of the context of choice in D5. For example, the salience to the bank teller of the combination to the safe is not harmfully distorted. However, the harm that may result from adverse features of the context of choice may occur in this type of case. If the key to the back door of the bank mentioned above is more salient to the agent than it should be, then this may exert a harmful influence on her since she may be more likely to choose to carry it about with her. Furthermore, this effect may intersect with her weak character, if she has one.

1.7.6 "Morality and Choice"

The harm of "morality and choice", D6 (Dworkin, G., 1988, p.75), results from an additional option that is selectively available to some agents in a society that does not generally value this option. The additional option harms society as a whole and not individual agents in virtue of its availability. Thus the harm of D6 is present before any choice is made by the agent. This means, first, that the character of the agent can have no bearing on D6, and second, that there are no adverse features of the context of choice in D6. So D6 is a distinct harm from the two additional ones that I defend.

1.7.7 "Paternalism and Choice"

I noted above that the harm in "paternalism and choice", D7 (Dworkin, G., 1988, p.76), appeals to aspects of the agent's character. By way of a reminder, an agent in D7 recognises in advance that she may be harmed if she takes up an additional option and then contracts with other people to restrict her access to that option. In this harm, then, the agent recognises that she has a weak character of the relevant type and takes action to prevent herself from exercising this weakness. The harm in D7 thereby confirms the relevance of weak character to harm in certain choice situations. However, Dworkin describes a sub-type of D7 in which the agent is not "tempted" to make the relevant choice but instead makes it by mistake. So Dworkin says that he "would not want to have a bomb connected to a number he could dial on his phone, because [he] might dial it by mistake" (Dworkin, G., 1988, p.76). So the sub-type of D7 does not depend on the agent's character, but is it possible to identify differences between the main harm in D7 and harms due to the types of weak character that I defend?

First, the agent in D7 is aware of the type of weak character in herself that is relevant to her choice situation, but this cannot be said of all people with this type of weak character. So it is possible for an agent with weak decision-making skills to be unaware of this weakness. As a result of this lack of awareness, an agent with weak decision-making skills would not feel any need to contract with other people to restrict her access to the harmful option. Furthermore, the other two types of weak character that I identify, namely undue self-deprecation and undue lack of confidence in judgements, do not influence the agent to be tempted by the new option against their better judgement. The new option gives the agent the opportunity to exercise her weak character because she merely feels that it is suitable for her, and she has no reason to put preventative measures in place. So the harms due to weak character that I defend are broader in scope than the type of weak character in D7. I now turn to the role in D7 of my second harm, which results from adverse aspects of the context of choice.

The salience to the agent of the additional option need not be harmfully distorted in order for harm to result to the agent in D7. However, as in previous cases, if the additional option is more salient to the agent than it should be, then this will result in harm to her because she is more likely to select it. The harm to the agent in this case may also intersect with weak character: there is a greater chance that the agent will exercise her weak character if the additional option is more salient to her.

1.7.8 Preference for the Default

In Velleman's first harm, preference for the default, V1 (Velleman, 2015, p.10), the agent and the people offering the new option differ in their valuation of the additional option (an invitation to dinner is the additional option in one of the cases used by Velleman to illustrate this harm). The agent does not value the new option before it is offered, and would prefer the default situation, but the people making the offer do value it. If we consider, first, an agent who does not have a weak character we can see that the offer to her of the new option is unwelcome and may be harmful. The offer is harmful since she now has to actively choose what was previously the default option. A further source of harm in this situation is that the agent must also inform the person making the offer that she does not want to accept it. Furthermore, if the agent were to be re-offered the same option it can be seen that the harm may be cumulative. Last, the more significant the additional option is in the eyes of the agent and the person making the offer, the more likely the agent is to be harmed. Consider, for example, an offer to an agent of a well-paid job in a field relevant to her skills but in a distant geographical region. There are high stakes in this offer, perhaps, for example, because the agent currently has a poorly paid job, but also has family commitments where she currently lives, and this may make it especially difficult and harmful for her to turn it down. So having a weak character is not necessary for harm to result in V1.

If, however, an agent in the type of case used to illustrate V1 has an undue lack of confidence in her judgements based on her values, she may be at increased risk of harm. In the example above, the agent may, first, struggle to justify to herself her decision to reject the offer if she unduly lacks confidence in her judgement relating to this. Second, she may also face an increased difficulty in revealing her choice to the person who has made the offer if she is unsure of her justification for the choice. So the harm that accrues to an agent with a weak character in this type of case intersects with the harm identified by Velleman in V1. If the agent has a weak character in this type of choice situation, then the harm is exaggerated when compared with the harm that may accrue to a robust agent.

An exaggeration of the harm in cases used to illustrate V1 may also result if there are adverse effects of the context of choice. If an agent in this type of case finds certain aspects of the choice situation more salient than she should do, then this could result in her being more harmed. She may, for example, find the disparity between her values and the values of the person offering the option more salient than she should do, and

this will result in her finding it harder to justify her choice and inform the person making the offer. However, the agent could still be harmed in this type of case if there was not this adverse feature of the context of choice.

1.7.9 Harmful Implications

Finally in my revisiting of Dworkin's and Velleman's harms I turn to Velleman's second harm, "harmful implications", V2 (Velleman, 2015, p.11). This is the harm to an agent that results from what the option itself expresses about her. So in Velleman's example, the offer of remedial assistance by a lecturer to a student expresses the view that the student is underperforming in her studies despite her performing to an adequate level. A student should rightly be concerned at the implication that she is underperforming. Furthermore, she should be concerned and harmed at this negative implication even if she has a robust character. An agent may be more likely to be harmed in this type of situation if the option being offered is a graver one—consider for example an offer of a termination of pregnancy for a child with Down's syndrome. The implication in this case is that the disorder is so significant that allowing the pregnancy to go to term is questioned. The harm in this type of case may also be cumulative if the harmful implication is repeated.

The agent will also be susceptible to being harmed in this type of choice situation if she either has a weak character or there are adverse features of the context of choice. If the agent has an undue lack of confidence in her abilities then she will have more reason to be harmed by the negative implications contained in the offer than if she is more robust in her character. So even if the agent has no need of the additional option, her undue lack of confidence in her needs will lead her to believe that she does need it. A similar harmful effect will occur if the agent finds the new option and its harmful implication more salient than she should do. One possible mechanism for this change in salience having a harmful effect could be if the authority of the person making the offer is more salient than it should be. For example, an offer of remedial help from a person who is perceived to have more authority will have more impact than an offer from a person who is perceived to have less authority.

1.8 Conclusion

This chapter has argued that certain types of weak character and what I term the "context of choice" may give rise to two harms to an agent who is presented with an

additional option. First, harm may be accrued by an agent who is given the opportunity to exercise any of three types of weak character, and second, harm may be accrued by her if the context of choice harmfully alters the salience to her of certain aspects of the choice situation. These claims go beyond those of Dworkin, G. (1988) and Velleman (2015) who argue that an agent may potentially be harmed in several different ways if she is offered an additional option. An analysis of their nine harms reveals that each of them is different from the two additional harms that I identify. They are different because it is not a necessary feature of Dworkin's and Velleman's harms either that the agent has a weak character of the three types I defend or that the context of choice is adversely configured. However, both of my additional harms may additionally occur in the types of case described by Dworkin and Velleman.

Having defended arguments that an agent may potentially be harmed if she is presented with an additional option, and that this harm is more likely if she has any of three types of weak character, it is now necessary in chapters two and three for me to defend full accounts of character and weak character.

Chapter 2. An Account of Character

2.1 Abstract

In this chapter I outline a foundational account of character from Aristotle in his Nicomachean Ethics (NE). This is a necessary first step in my line of argument that certain key types of weak character may increase the chance of an agent being harmed if she is offered an additional option. Aristotle's account of character is a suitable one for my purposes because it can provide the necessary resources for an explanation of the three types of weak character that I identify in chapter three. Aristotle's account of character includes a fundamental role for the emotions; these are given a low priority in Kant's account so, second, I defend Aristotle's account of character against Kant's. Third, I compare selected modern accounts of character with Aristotle's one, and defend a claim that these do not provide a significant advance on Aristotle. Fourth, there are objections to the notion of character from Situationist ethics. However, these are based on empirical research which has flaws and does not take proper account of character-based explanations for their findings. In the remainder of the chapter, it is necessary for me to make plausible a role for the notion of character in ethical theory. This is necessary because the force of a claim that weak character may increase the chance of an agent being harmed if she is offered an additional option will depend on this role for character in ethical theory.

2.2 Introduction

In chapter one, I defended an important role for the notion of character, in particular weak character[1], in explaining potential harms to an agent who is presented with an additional option. The account of character I used in chapter one was a preliminary one, so in chapter two I will defend a more complete account. This will then enable me explain three key types of weak character in chapter three.

First, I outline a foundational account of character from Aristotle in the Nicomachean Ethics (NE), (2002). I focus in this account on the important concepts of habituation (*ethismos*), disposition (*hexis*), choice (*prohairesis*), the mean (*meson*), rationality (*orthos logos*) and practical wisdom (*phronesis*) and differentiate Aristotle's six types

[1] Together with a role for the context of choice.

of character, showing how they are relevant both to an agent's behaviour and to a moral evaluation of the agent herself.

Emotions also have an important role in Aristotle's account, and this fundamental role is missing in Kant (Beiser, 2007). So, second, I defend Aristotle's account of character against that of Kant. Kant's account of character makes it impossible for an agent to know if she is acting virtuously if she derives pleasure in doing so. Furthermore, Kant demotes a role for emotion in character to the so-called "sensible" character, where it is subordinate to the "intelligible" character (Athanassoulis, 2005). Both these features of emotion in Kant's account underutilise a fundamental element of human psychology, and as a result, Kant's account is less well equipped than Aristotle's to explain weak types of character.

Third in this chapter, I make plausible a claim that contemporary accounts of character from Kupperman (1991), Goldie (2004) and Williams (1973) do not provide a significant advance on Aristotle. Their accounts have features such as constitutive psychological features that are shaped into character and dispositions that form over time, but these are also present in Aristotle's account. Aristotle, unlike Goldie, does not make use of the notion of life as a narrative. However, I argue after Williams (2009) that this potential aspect of character has limited application for agents in the course of their lives.

Fourth, I defend Aristotle's account of character against objections that have been raised from the perspective of Situationist ethics, which claims that it is situational features alone that settle how an agent behaves, rather than character, e.g. Harman, G. (2000) and Doris (2002). The interpretations of empirical studies that are the basis for Situationist claims are flawed since they fail to take account of Aristotelian sub-types of character (Athanassoulis, 2000). Furthermore, the design of some of these studies is also flawed since, first, they do not test repeated instances of behaviours (e.g. Annas, 2005), and second, they also pit different virtues against one another thereby posing more of a challenge to subjects than if they pitted vices against their corresponding virtues (Kristjánsson, 2008).

Last in this chapter it is necessary for me to defend a claim that the concept of character has a central theoretical role in an account of good ethical judgement. Without this central role I cannot defend a place for weakness of character in explicating potential

harms to an agent who is offered an additional option. Character's role is that it offers the possibility of a deeper moral evaluation of an agent based on an understanding of the origins, development and responsibility for her action, and on an understanding of the agent's psychological state, including psychological tensions, whilst she acts.

2.3 Aristotle's Overall Project in the Nicomachean Ethics and his Theory of Character

Aristotle's account of character forms part of his overall project, which has three key parts, in the Nicomachean Ethics (NE). First, he aims to provide an argument in defence of a particular conception of the ultimate good for man (*eudaimonia*), second, he gives an account of virtue (*ethike arête*), and third, he seeks to explain the relation between virtue and the ultimate human good. For reasons that are not relevant to my purposes, which are derived from a claimed defining function for humans ("activity of soul and actions accompanied by reason" (NE 1098a14)), Aristotle holds that *eudaimonia* is an "activity of soul in accordance with excellence […] in a complete life" (NE 1098a16)[2]. Relevantly for my purposes, however, he claims in the Eudemian Ethics (EE 1220a39-b3) that dispositions of the soul acquired through habituation are the *character* of the agent (Aristotle, 1992), and as I will show, virtue is one of his six subtypes of character. So, first in this section, it is necessary for me to briefly lay out Aristotle's account of the soul, so that we can start to see what character is. Second, I will explain Aristotle's account of how humans have natural psychological tendencies, and the different ways in which these may be habituated (*ethismos*). In Aristotle's account, habituation of different tendencies, whether or not with guidance from another person, results in the relevant part of the agent's soul developing a more settled state (*hexis*); in Aristotle's account this is how an agent's character is formed. An agent's character results in her acting in certain ways depending on her choice (*prohairesis*) and emotions (*pathe*), so I will also explain the relevance of choice and emotions to character in Aristotle's account. An important feature of choice in the NE account is that it is voluntary, since things that are voluntary open the agent to praise or blame, so I will also explain Aristotle's account of the voluntary (*to hekousion*).

---

[2] I identify citations and quotes from the Nicomachean Ethics (NE) using Bekker numbers so that they can be located in any edition of the NE.

Having explained character, i.e. how it is developed, what it is, and how it is revealed by choice, emotion and action, I give in the last section an account of Aristotle's six different types of character. These types range from the excellent to the imperfect. I discount one of these types—superhuman excellence—because of its similarity to one of the other types—excellence of character—and, following Aristotle, differentiate between the remaining five character types on the basis of the respective roles in each of them of choice and emotion in determining action. Aristotle further claims that excellent character aims at a "mean", so an explanation of the mean will be necessary to mark off this type of character from the other types. An account of Aristotle's six types of character is important for my purposes in chapter three where I defend an account of weak character.

2.3.1 Constituent Parts of the Soul

As I have said, character for Aristotle is a particular state (*hexis*) of the soul. So first, I briefly describe one aspect of his account of the soul in order to set the scene for how this state may arise. Aristotle claims that "[t]here are three kinds of things in the soul—feelings (*pathe*), capacities (*dunameis*), and dispositions (*hexeis*)" (NE 1105b19-20). In his account, the various character states, including virtue, are dispositions of the soul, and he thereby contrasts dispositions with feelings and capacities: "as for dispositions, it is in terms of these that we are well or badly disposed in relation to the affections, as for example in relation to becoming angry, if we are violently or sluggishly disposed, we are badly disposed, and if in an intermediate way, we are well disposed" (NE 1105b25-27). This passage brings out an important relationship between dispositions and affections (or emotions). Emotions, and how we are disposed in relation to them, are an important aspect of character so, below, in the section on Aristotle's different types of character, I will show how they help to identify character types. In the next section, however, I start with an explanation of the origins of an agent's character.

2.3.2 The Natural Tendency of Humans to Experience Emotions

A starting point in Aristotle's explanation of how humans acquire character is their natural tendency to experience certain feelings, such as "appetite, anger, fear, boldness, grudging, ill will, friendly feeling, hatred, longing, envy, pity—generally feelings attended with pleasure and pain" (NE 1105b20-23). We have feelings in relation to most aspects of our lives. For example, we may experience pleasure during a walk in the countryside on a sunny day in the company of friends. Importantly, we may *act*

from emotion (a characteristic that adult humans share with children and non-human animals). Furthermore, according to Aristotle, our emotions are directed at certain occurrences. As Hursthouse says, after Aristotle: "we come into the world, for the most part, set up to enjoy and be distressed by, broadly speaking, some of the right things: for example, eating, being liked or loved, and others' enjoyment, on the one hand, and physical damage, being thwarted, and others' distress or anger, on the other" (Hursthouse, 2006b, p.111). These natural dispositions are the origins of character; they are an important aspect of the material which is shaped into character. But what is the method by which this starting material is shaped into character?

2.3.3 Emotions and Habituation

An agent will tend to repeat behaviours—whether they are right or wrong—in certain situations when the behaviour itself is associated with pleasant feelings[3]. An agent may also experience pleasure associated with certain behaviours as a result of approval from other people. The more that behaviours associated with pleasant feelings are repeated, the more an agent in the same situation in the future will tend to behave in the same way, and to experience the same emotions. By contrast, an agent will tend *not* to repeat behaviours associated with painful feelings, including shame, disapproval or anger from others; she will become habituated to behave in the same way *less often* in future (NE 1128b15-18; 1179b8ff). Charles helpfully identifies the key passages relevant to habituation in the NE: an agent should have "early training and habituation to feel pleasure and pain as one should (1104b12-13; 1105a4f.), and punishment when one fails (1104b16-18; 1179b24-29; 1180a5; 1105a5ff.), reproach, exhortation (1102b34f.) and the efficacy of shame (1128b15-18; 1179b8ff.) in regarding certain desires as ignoble (either because excessive within the permissible range or because outside the permissible range)" (Charles, 1984, p.180).

This process of habituation is the process by which character forms. And character itself, as the product of habituation, is the resultant disposition or state of the soul (Sherman, 1989). I will say more about habituation below in relation to the specific character types, since habituation under different conditions helps to determine the types of character that an agent may acquire.

---

[3] An agent's *choice*, as we shall see in a following section, also has an important role in respect of her tendency to repeat certain behaviours.

2.3.4 Habituation and Stability of Character

I have briefly explained how, in Aristotle's account, character emerges as a result of habituation. Whichever way an agent is habituated to feel and behave, these emotions and behaviours will tend to become more stable the more they are practised. However, the notion of stability does not necessarily mean that the agent will always behave in the same way when she is in similar situations; stability should be taken to mean that it is more likely than not that the agent will behave in the same way in a similar situation, but it is also possible that she could choose to behave in a different way—this is the way in which an agent's character may evolve over time under her own control. Aristotle indicates this progression to a more stable state in two of the words he uses to describe character. As Annas observes, Aristotle uses both *hexis* and *diathesis* to describes states of the soul: "A *hexis* differs from a *diathesis* by being more stable and more long lasting. Such are the different types of knowledge and the virtues" (Aristotle Cat. 8.8b27-29, in Annas (1993), p.50). So, at an earlier stage in character development the relevant part of an agent's soul may be a *diathesis* and later, when character is more fully formed, it may be a *hexis*. Aristotle describes a further normative aspect of *hexis* in the *Metaphysics*: "A *hexis* is a disposition according to which that which is disposed is either well or badly disposed, either in itself or in relation to something else, for example, health is a *hexis*, for it is such a disposition" (Met. 5.20 1022b10-12, quoted in Achtenberg (2002), p.111). As I have said, *hexis* in the context of character (rather than health) is a good or bad disposition in respect of affective states. Importantly in Aristotle's account, the agent may choose her actions, so I now turn to Aristotle's account of choice and how this relates to character.

2.3.5 Aristotle's Account of Choice (*Prohairesis*)

In Aristotle's account, choice (*prohairesis*) has a fundamental role, with emotion, in determining how an agent behaves in situations that she may encounter. An agent is able to choose what to do, and may consequently repeat and habituate certain behaviours with their associated emotions. As a result of these choices, her character is shaped. Conversely, an agent may choose not to repeat certain behaviours in favour of other ones. So in the next section I explain Aristotle's account of *prohairesis* including its role in his analysis of character. Following my explanation of Aristotle's accounts of habituation, dispositions, and choice, I will be in a position to describe his six different character types.

Aristotle's account of *prohairesis* places it in the realm of what is voluntary (*to hekousion*) for humans. That is to say that, for Aristotle, *prohairesis* must meet two conditions: these are, first, that it has its source (*arche*) in the agent, and second, that it can only occur in an agent who also satisfies certain knowledge conditions. Furthermore, *prohairesis* follows a process of deliberation: *a prohairesis* is a decision about how to act to achieve an end that an agent wishes for. In this section I outline Aristotle's account of what is voluntary, differentiating the components of voluntariness, and then focus on the part of the voluntary that is *prohairesis*, or "deliberated choice". After describing the mental processes that precede *prohairesis*, I defend a claim that Aristotle's account of *prohairesis* can explain human decision-making. Furthermore, I argue that *prohairesis* can add precision in the differentiation of character types that is not possible if mere actions are analysed.

2.3.5.1 *Prohairesis* and the Voluntary

Aristotle's account of *prohairesis*, in book III of the NE, starts with an account of things that are voluntary and involuntary for an agent. The importance to him of the distinction between the voluntary and involuntary is that reactive attitudes such as "praise and blame" (NE 1109b31) are appropriate for what is voluntary. So unless he is able to establish that *prohairesis* is voluntary, it will not be possible to adopt appropriate reactive attitudes to an agent on the basis of her choices. Moreover, in the case of involuntary actions a different set of reactive attitudes, such as pity, may be appropriate. Aristotle makes two key claims about things that are voluntary: first, that they have their origin inside the agent, and second the agent must satisfy certain knowledge conditions.

2.3.5.2 Voluntariness

In Aristotle's account, the first condition for voluntariness of an action is that it originates in the agent: "a person acts voluntarily in the cases in question; for in fact in actions of this sort the origin of his moving the instrumental parts is in himself, and if the origin of something is in himself, it depends on himself whether he does that thing or not" (NE 1110a15-18). The corollary of this is that involuntary things have their origin outside the agent, "such that the person [...] contributes nothing, as for example if a wind were to carry him somewhere[.]" (NE 1110a3-4). However, Aristotle identifies some actions that have an internal source as being mixed cases, i.e. there are features of them that suggest that they are both voluntary and involuntary.

Aristotle's examples of mixed cases are, first, a tyrant who orders someone to do something shameful under threat to the lives of his parents and children, and second, a sailor who throws his cargo overboard to protect his boat in a storm (NE 1110a5-12). The two features that suggest that these actions are voluntary are that they are chosen by the agent, who can either do them or not do them, and that they arise within the agent. The feature that suggests that the actions are involuntary is that "no one would choose anything of this sort for itself" (NE 1110a19-20). As Charles says: "mixed actions are 'mixed' precisely because two conditions conflict in their case: (a) they are chosen in certain conditions as means to a further goal, and (b) they are not wanted (for their own sake)" (Charles, 1984, p.60). So in other words, the agents in the examples have the option not to commit either act, but choose to do so only as a means to a further goal—namely safety of either family or crew.

It is illuminating at this point to consider appropriate reactive attitudes in relation to the two agents in Aristotle's examples. Both of the agent's in Aristotle's examples could be praised for their actions despite each of the acts being one that would not be chosen "for itself". The corollary of this is that the agents would be blamed if, for example, they had allowed either family or crew to come to harm when the option of preventing this was available. The "for itself" stipulation about what the agents should choose is an empty one in each context, since there is no possibility for them either that the tyrant will not threaten the family or that the storm will abate. So it can be seen that Aristotle's mixed cases are in fact examples of voluntary behaviour, but that the behaviour is under constrained circumstances. An important feature of both these examples is that the agent appears to have knowledge of what may happen in the case of refusing, or not refusing, to bow to the tyrant's threats, or not throwing, or throwing, the cargo overboard. This feature raises the second of Aristotle's conditions for voluntary acts, which is that the agent satisfies certain knowledge requirements.

Aristotle's knowledge based condition for voluntary action has two parts, and this division into two parts is possible because he has divided up rational motivation into "what the agent does (the action) [… and] the goal for the sake of which the he does it (the good)—hence the distinction between the action and the *prohairesis* on which it is done" (Sauvé Meyer, 2006, p.150). So the agent has two different types of knowledge in voluntary actions. The flip side of this is that an agent can either be ignorant about the particulars of the action, or about the goals she is pursuing and whether they are good or bad ones. A (modernised) example of the first type of ignorance could be the

agent not knowing whether or not there was a person behind the bathroom door that they were shooting through, or not knowing whether or not the gun contained blanks or live ammunition (NE 1111a3-15). In both these cases the action is involuntary. However, the second type of ignorance, demonstrated in cases where the agent is unsure about what is a good goal to have, is voluntary.

The voluntary, for Aristotle, is shared by adults, children and non-human animals. Only adults, however, can act voluntarily from each of appetite (*epithumia*), temper (*thumos*) and wish (*bouleusis*). Children and non-human animals are capable of acting voluntarily only from appetite. The first two of these "three types of desire (*orexis*)" are immediate impulses: "appetite, [...] towards the pleasant or away from the painful [,..] and temper [...] in response to aggression or insult. Wish [, however, is] for something judged good and not immediately attainable" (Broadie, 2002, p.314). Broadie presumably thinks that the *epithumetic* as well as *thumetic* are both impulses because they leave no time for a decision. In contrast, wishing or desiring an end, because the end is not immediately attainable, requires a reasoned decision or choice (*prohairesis*) as part of a process of attaining that end: "wish is more for the end and decision is about what forwards the end" (NE 1111b27). It should be noted that there is no need to wish for an end that is immediately attainable; in this case, if possible, the agent merely achieves the end. So decision, or *prohairesis*, then, also emerges from this passage in the NE as a distinct part of the voluntary.

2.3.5.3 *Prohairesis* and Deliberation

If the internal source and knowledge conditions described above are satisfied, then this meets one of the conditions for an agent to be able to make her reasoned choice, her *prohairesis*. In order to form *prohairesis*, however, the agent must also deliberate towards an end. And in order to deliberate towards an end the agent must form, first, a conception of what is valuable and thereby worth pursuing, and second, desires or wishes that match those ends. Clearly, an agent cannot reasonably desire anything, as in the idiom "wishing for the moon". However, other desires can be accepted as more reasonable in the context of the agent's life. And these can be termed "preferential desires": ones that are "formed in the light of deliberation as to how to attain goals, where those goals reflect a conception of the good" (Megone, 1998, p.221).

Consider an agent who wishes or desires a tasty and healthy meal in the evening. This desire is a reasonable one for most people (but perhaps not for people starving under

siege in Syria). Once the desire is formed, the agent must then deliberate about the different ways of achieving the desired end. In this case, the deliberative process will involve different types of food with their associated costs, and different ways of procuring and preparing them. Note, however, that Aristotle specifies that deliberation is not about particulars, "e.g. about whether this is a loaf, or whether it has been cooked as it should; for these belong to the sphere of perception" (NE 1113a1-3). However, deliberation does make use of particular premises which may be generated through perception.

It might seem that deliberation and desire are not inter-related with each other in a dynamic way that permits each of them to influence the other—that deliberation just takes over the baton of mental activity from desiring. However, if we consider a food based example from Sherman, it can be seen that the relationship between desire and deliberation is a dynamic one. In her example, Sherman desires "truffles with wild duckling for supper". Since, on deliberating, she realises that both these foodstuffs are hard to obtain, it can be seen that "deliberation will often provide a test of the practicality of desires" (Sherman, 1989, pp.65-66). The process of deliberation has to reach a conclusion so that the agent can proceed towards action, because "if a person deliberates at every point, he will go on for ever" (NE 1113a3). The end point of a process of deliberation is *prohairesis* (choice), or decision, so

> "[w]hat we deliberate about and what we decide on are the same, except that what is decided on is, as such, something definite; for it is what has been selected as a result of deliberation that is "decided on". For each person ceases to investigate how he will act, at whatever moment he brings the origin of the action back to himself, and to the leading part of himself; for this is that part that decides" (NE 1113a4-9).

Sherman, on reasoning that truffles with wild duckling is not a practical meal, might have ended a process of deliberating about what to eat for supper with a decision to buy the ingredients for, and cook, a Mediterranean omelette; eggs and mushrooms are much easier to find, cheaper to buy, and probably just as healthy as truffles and duckling. Her decision in this case would be consistent with Aristotle's claim that "[w]hat we do deliberate about are the things that depend on us and are doable" (NE 1112a31). One possible objection here is that truffles and wild duckling are merely particulars and are not an end; the proper end in this case should be the more general one of health. So perhaps truffles and duckling are an *intermediate* end, and consequently a more practical end for deliberation than health itself. An agent who deliberates about health

as a broader end will not only deliberate about food but about many other inter-related aspects of her lifestyle, including contact with other people, exercise, work, and sleep[4].

So far in this section I have shown that *prohairesis* in Aristotle's account is a sub-class of the voluntary—i.e. that it has its origin in an agent who also has knowledge both about the nature of the action she is choosing and a conception of what the goal of the action is. Furthermore, *prohairesis* is the end point of a process of deliberation, and deliberation, in turn, is inter-related with what the agent desires. However, since *prohairesis* has a fundamental role in Aristotle's account of character, it is necessary to show that *prohairesis* does not relate to mere technical areas of decision-making. This part of an account of *prohairesis* is important because Aristotle has been accused by Sherman (1989) of using examples of technical decision-making in his account of *prohairesis* in the NE. So I now defend her claim that *prohairesis* is applicable to character and not applicable in the sphere of the technical.

Sherman summarises her argument about the applicability to character of *prohairesis* in Aristotle's account by saying that his "examples of deliberation do not reveal the complexity of acting from character" (Sherman, 1989, p.76). His examples of deliberation (and decision), in contrast are relatively simple technical ones involving agents from different professions:

> "We deliberate not about ends, but about the things that promote the ends. For neither the doctor deliberates if he should heal, not the orator if he should persuade, not the politician if he should produce good order, nor does anyone else deliberate about his end. But positing the end, they consider how and through what means it will be achieved. And if it seems that it can be achieved by several means, they consider further by which one it is most easily and best realised. And if it is achieved by only one means, they consider how it is achieved by that means, and how that means be discovered… And if they come upon an impossibility, they give up the search, e.g. If they need money and this cannot be secured; but if a thing appears possible, then they try to do it" (NE 1112b12-27).

In deciding how to act, each of the three professionals in the passage above can be seen to be following a simple syllogism consisting in a major premise, respectively to heal, persuade, and produce good order; a minor premise arrived at through deliberation and decision (or even by looking in a textbook), e.g. treatment x heals the disease in question (in the case of the doctor) and an action, to give the patient treatment x. It could however be objected, for example in the case of the doctor, that Sherman has not

---

[4] I discuss deliberation about ends and means in more detail in chapter three, in the section on the intellectual virtue.

made a clear distinction between the technical and ethical. Consider a vicious doctor who does not merely fail to heal, but who fails to heal whilst having the intention of making money[5]. In reply, the doctor in this case has two goals which are, first, healing—a technical one in which he fails (perhaps deliberately, in order to harm)—and a second goal of making money—an (un)ethical goal. If we look at Aristotle's passage above, we can see that he does not elaborate the technical aspect of medicine with an ethical dimension: the doctor fails in the technical practice of medicine, presumably without any intention to harm. So Sherman's criticism seems well founded. It is not difficult, however, to move beyond technical applications of the practical syllogism to examples which have relevance to moral reasoning.

Sherman gives an example of deliberation related to medicine which is not merely technical and which reveals moral reasoning. In the process of choosing to become a doctor an agent's decision "may result as a choice about how best to earn a living in a way that is at once socially prestigious and humanitarian" (Sherman, 1989, p.71). In this example the agent does not merely give a drug which is medically specified as a treatment in the minor premise of a practical syllogism and which satisfies the requirements of a major premise, but instead deliberates and decides on a complex course of action that may also, as I describe below, reveal aspects of the agent's character.

A second claim made by Sherman (1989) about *prohairesis* is that it can apply to actions that that are not merely proximate in time. Her motivation for making this claim is that one of Aristotle's examples of a syllogism implies a series of actions that occur in sequence, with the final action being in the future. Sherman (1989, p.69) cites a syllogism in the *De Motu* (*MA* 698a4):

1. I need a covering.
2. A coat is a covering.
3. I need a coat.
4. What I need I must make.
5. I need a coat.
6. I must make a coat.

---

[5] We have to assume in this case that the failure to heal, perhaps by not being readily apparent to the patient, results in monetary gains rather than losses.

7. And the conclusion that I must make a coat is an action.
8. And he acts from a starting point.
9. If this is to be a coat, first there must be this.
10. And if this, this.
11. And he does this straightaway.

This syllogism consists in a series of actions that must occur in a certain order—"this. and if this, this."—which culminate in achieving the specific goal of having a coat. The actions that result in the goal of having a coat occur in a specific sequence and are thereby separated in time. So the first (proximate) choice, "this", is the first in a series of non-proximate "this" choices.

In contrast, the examples of deliberation and choice given by Aristotle in the NE issue in immediate action, such as persuasion in the case of the orator. However, for example, both the politician and doctor may also deliberate and decide about actions which take place in the future. A doctor may set into motion a complex and lengthy plan of treatments—perhaps involving courses of chemotherapy and radiotherapy—and the politician may plan for future actions in the same way. In the case of an individual agent deliberating and deciding about the future course of her life, it follows that she can choose to perform a sequence of actions in ways that promote her interests into the future. As Sherman says, "if one as a rational agent is to be more than a bundle of disparate streams of interests, then part of planning will involve the coherence of ends side by side (synchronously) over time (diachronously) and the promotion of actions in the light of that pattern of ends. [...] The general point is that one's character is integrated and stable to the extent to which one can form systematically related intentions that realise one's general ends" (Sherman, 1989, p.76)[6]. So, if the individual above who decides to become a doctor so that she can enter a profession that is both "socially prestigious and humanitarian" is growing up, say, in Inverness, she may also choose to study Gaelic in order to practice in the Highlands and Islands of Scotland.

In the previous section, following Sherman, I have defended Aristotle's view of *prohairesis* as being a deliberated choice to act or embark on a sequence of actions that

---

[6] An alternative explanation of deliberation about ends is that ends are the goals of the practically wise person. In Sherman's account it is *prohairetic* deliberation that resolves this rather than rational reflection on ends. As in the footnote above, I say more about practical wisdom in the section in chapter three on intellectual wisdom.

is not technical; *prohairesis* is a means to achieving a practical, desired, end. Furthermore, I have argued that *prohairesis* applies to complex acts relevant to the way an agent leads her life; she deliberates on multiple coherent ends and the means to achieving them. How an agent chooses to lead her life might thereby be expected to be revealing about her character and I now explain the role of *prohairesis* in Aristotle's six character types. In this section I will also argue that *prohairesis* is more revealing of character than mere actions.

2.4 Aristotle's Six Character Types

Aristotle's definition of virtue, one of his six sub-types of character, is that it is

> "a disposition issuing in decisions, depending on intermediacy of the kind relative to us, this being determined by a rational prescription and in the way in which the wise person would determine it. And it is intermediacy between two bad states, one involving excess, the other involving deficiency; and also because one set of bad states is deficient, the other excessive in relation to what is required both in affectations and in actions, whereas excellence both finds and chooses the intermediate" (NE 1106b36-1107a5).

We can abstract from this definition a paraphrased definition of character, which is that it is a disposition to generate decisions to act and to act based on reasoning and emotional states. Aristotle elaborates his account of character by identifying six different character types which exhibit different moral qualities in the agent. There are three unwanted states of character, which are being bad, lacking self-control and being a brute (NE 1145a16-17), and the contraries of two of these are excellence of character (virtue) and self-control. The contrary of the third character state is "superhuman excellence", which has qualities "of a heroic or even divine sort" (NE 1145a21). The link between character and *prohairesis* in Aristotle's account is that the choice the agent makes after deliberating on the best means to achieving a desired end contributes—with her emotions—to revealing which of the six different character types she has. A further necessary contribution to revealing the agent's character type is made by her actual behaviour after making a choice; that is to say that the agent's character is shown both by her choice of action and by her actual behaviour. For example, the agent's action may not be aligned with her *prohairesis*, but may instead be in line with outlier emotions. I now explain Aristotle's six character types, but first, it is necessary for me to say more about the role of emotion in determining behaviour, because in Aristotle's account, as I have said, reason and emotion both have roles in determining the agent's behaviour.

2.4.1 Emotions and Character

The emotions (*pathe*) relisted here, such as "appetite, anger, fear, boldness, grudging ill will, joy, friendly feeling, hatred, longing, envy, pity—generally, feelings attended by pleasure and pain" (NE 1105b21), have a double role in determining behaviour. First, they help to determine what is morally relevant through their role in perception: "[t]hrough the emotions we come to recognise what is ethically salient" (Sherman, 1989, p.38)). Second, the emotions can motivate actions and, as above, in Aristotle's account this is a feature that is shared by non-human animals, children and adults (NE 1111b14). However, the emotions that motivate actions in children and non-human animals are not ones that are responsive to what is ethically salient. So an agent (or animal or child) may experience fear when she sees a fire that appears to be growing out of control, and may then be moved to run away from the fire as a result of her fear. However, only the adult in this case will be responding according to what is ethically salient, and the adult's action (with her choice) will be revealing of her character. So "[w]henever one acts in a way that displays character, Aristotle believes, one will be manifesting one or another of these and similar emotions" (Urmson, 1973, p.224). However, alongside emotion, reason (or choice) also has a role in generating both decisions to act and actions themselves; reason, emotion and action together are revealing of the agent's character (NE 1106b36-1107a5). So in the next section I show how in Aristotle's account these three aspects of an agent reveal her character type.

2.4.2 Super-human and Excellent Character

First in Aristotle's taxonomy of character types are super-human and excellent character. Agents with either of these two character types are stably disposed to deliberate and decide to act in ways that are excellent or virtuous. Requoted from above, they have

> "a disposition issuing in decisions, depending on intermediacy of the kind relative to us, this being determined by rational prescription and in the way in which the wise person would determine it. And it is intermediacy between two bad states, one involving excess, the other involving deficiency; and also because one set of bad states is deficient, the other excessive in relation to what is required both in affections and in actions, whereas excellence both finds and chooses the intermediate" (NE 1106b36-1107a6).

In other words, if an agent with a super-human (*theios*) or excellent (*arête*) character is making a decision to act in circumstances that may, for example, be frightening, such as in proximity to a fire, she will act courageously: her *prohairesis* is intermediate relative to her, and is also in line with the wise person's choice. Furthermore, her emotions and

actions are intermediate: she experiences the correct amount of fear—not too much or too little—and she neither necessarily runs from the fire through disproportionate fear, nor stands still through too little fear. Super-human and excellent agents are also stably disposed to derive pleasure from behaving in the right way (e.g. courageously) and do not experience internal conflict between the result of their rational decision and the affective states they experience. Excellence of character, then, is an "unconditional preparedness to act, feel, and in general respond in the ways typical of the humanly excellent person" (Broadie, 2002, p.19).

Excellent character has a special place in Aristotle's account because it is necessary for human flourishing (*eudaimonia*); the other character types that I will explain do not have this special status. There are features particular to the acquisition of excellent character, to which I now turn, that help to set it apart, and these will also be important in chapter three when I explain weak character (contrasted with robust character). In Aristotle's account, acquisition of excellent character is first differentiated from acquisition of intellectual excellence[7]. He then argues that acquisition of excellent character will require a process of habituation through performing virtuous acts.

Aristotle differentiates excellences of the intellectual and practical parts of the soul through an analysis of their origins and how they "increase"; the intellectual part "mostly comes into existence and increases as a result of teaching [...] whereas excellence of character results from habituation" (NE 1103a16). Furthermore, Aristotle claims, excellences of character can develop in us by habituation "because we are naturally able to receive them" (NE 1103a25). Aristotle illustrates this natural ability to receive excellences with a disanalogy between excellence of character and natural things. Natural things, he says, do not change with habituation; stones will always fall downwards no matter how often we throw them upwards. So, he says, excellence of character itself does not develop naturally. However, humans have a natural tendency to change with practice, so practice is an important component of acquiring an excellent character. There are, however, important features of habituation of character that mark it apart from practice.

Aristotle describes habituation in the following passage:

---

[7] I explain Aristotle's account of intellectual excellence in chapter three, in the section on lack of confidence.

> "we acquire the excellences through having first engaged in the activities, as is also the case with the various sorts of expert knowledge—for the way we learn the things we should do, knowing how to do them, is by doing them. For example, people become builders by building, and cithara-players by playing the cithara; so too, then, we become just by doing just things, moderate by doing moderate things, and courageous by doing courageous things" (NE 1103a31-b2).

In the following section he describes how an agent's emotions, which are a key component of character, may change with habituation, so that they are "intermediate" in relation to the sphere of the activity.

> "[I]t is through acting as we do in our dealings with human beings that some of us become just and others unjust, and through acting as we do in frightening situations, and through becoming habituated to fearing or being confident, that some of us become courageous and some of us cowardly. A similar thing holds, too, with situations relating to the appetites, and with those relating to temper: some people become moderate and mild-tempered, others self-indulgent and irascible, the one group as a result of behaving one way in such circumstances, the other as a result of behaving another way. We may sum up by saying just that dispositions come about from activities of a similar sort. (NE 1103b8-23)

So, in his account, Aristotle says that the ethical virtues come about through behaving in a virtuous way, and that in the process our emotional responses in specific situations are modified to come into line with those of a virtuous person (*phronimos*). Importantly in Aristotle's account, the way in which our emotions are modified is not through force—described by Hursthouse as "the horse-breaking account" (1988, p.210)—but by changing as a result of responding to reason: "the appetitive and generally desiring part [of the soul] does participate in [reason] in a way, i.e. in so far as it is capable of listening to it and obeying it" (NE 1102b31-32). If successful, the end point of this process is that the agent is able to feel emotions at the right time, in response to the things and people she should, for the right reasons, and "in the way [she] should" (NE 1106b22-23). Missing from this account of acquisition of virtue so far is, first, the role of guidance and, second, so-called "intermediacy".

In Aristotle's account, guidance is necessary in the acquisition of virtue because, first, humans cannot be virtuous without acquiring virtue through habituation[8], and second, the right reasons, emotions and actions for the agent to habituate in any particular situation have to be shown to them. In the case of emotions, humans may be pre-disposed to feel pleasure and pain at some of the right and wrong actions, but there are

[8] Aristotle allows in the NE for the phenomenon of "natural virtue", which a person may possess at birth, but this type of virtue does not allow for voluntariness in its acquisition and in the actions of the agent, and so cannot count as true virtue (NE 1114b11-14).

also areas of human activity in which pleasure and pain may not be correctly associated with what is the right or wrong thing to do; it is in these areas that a person will need guidance in order to habituate the right reasoning, emotions, and behaviour. Behaviours in the realm of temperance are one example where pleasure and pain are not naturally aligned with right and wrong acts, since temperance is concerned with moderation in bodily pleasures such as those associated with food, drink and sex. Despite a natural tendency for an agent to be pulled towards such pleasurable activities, an agent with excellent character in the sphere of temperance will neither over-indulge nor under-indulge in bodily pleasures. In Urmson's words: "[i]f you are temperate you will like abstaining from orgies" (Urmson, 1973, p224). So an agent who is developing the virtue of temperance will need to be guided—at first, perhaps, by parents—towards moderation in bodily pleasures. Later, the agent may be able to "guide" themselves:

> "[e]xcellence […] depends on us, and similarly badness as well. […I]f it depends on us to do fine things and shameful things, and similarly not to do them too, and this, it is agreed, is what it is to be, respectively, a good person and a bad one, then being decent people, and being worthless ones, will depend on us" (NE 1113b6-15).

So over time, with guidance and choice, and if the agent generally acts in the same way when the same circumstances present themselves, a disposition becomes "self-reinforcing and self-perpetuating" (Broadie, 2002, p.19). Habituation of character can, however, go wrong. If a person has habituated the wrong emotions (NE 1179b18-19) then they may not respond to persuasion, and may only respond to force (NE 1179b28-9; in Hursthouse (1988, p.214)).

2.4.3 The Mean

I indicated at the beginning of the previous section that intermediacy or "the mean" has a crucial role in identifying different character types, so I now turn to Aristotle's conception of the mean. The process of habituation that I have described does not give a complete indication of how an agent should behave in any given sphere of human activity; we are merely told that she should follow the guidance of people who have already acquired excellence of character, with the result that pleasures and pains experienced by the agent, and her choices, will become correctly aligned with the right sorts of acts. Aristotle's move here is to argue for behaviour that is "intermediate between what exceeds and what falls short". Furthermore, this intermediate is not the arithmetical mean but the mean with "reference to the object" (NE 1106a29-30).

It is argued by Brown (1997), however, that with "reference to the object" does not mean with reference to the agent. In the example of an athlete that Aristotle gives to illustrate the mean, the correct diet is assumed to be one that is between two extremes of excess and deficiency, and relative to the athlete himself. Milo is an experienced athlete who will require a larger diet (but still one than is not excessive or deficient for him) than an inexperienced athlete. Brown (1997) has questioned this interpretation on the basis that Aristotle does not suggest that the mean is relative to the agent in any of his discussion of individual virtues (such as courage and moderation). She also singles out in Aristotle's account "the centrality of the *phronimos* in the definition of excellence of character (*ethike arête*)" (Brown, 1997, p.81), which is described by Aristotle thus: "[e]xcellence, then, is a disposition issuing in decisions, depending on intermediacy of the kind relative to us, this being determined by rational prescription and *in the way in which the wise person would determine it*" (my italics, NE 1106b36). Brown takes this centrality of the *phronimos* as an indication that the correct outcome of the decision-making process is determined by the decision that the *phronimos* would make, which would presumably not vary for different agents, and would be in a mean position for humans collectively. However, there may be particular features of individuals—such as Milo—that are different from features of the *phronimos* and which bear on what is right for them to do. Perhaps the *phronimos* in each particular case should be a case-sensitive one, e.g. a Milo-*phronimos*. An alternative way of putting this would be to say that the *phronimos* is able to take into consideration relevant aspects of the person. In this case there would be no need for a Milo-*phronimos.*

Further illumination of the doctrine of the mean comes through a discussion of individual behaviours such as temperate behaviour. Temperate behaviour will fall in a mean between two extremes which are self-indulgence and abstemiousness (or austerity—Aristotle uses the word "insensate" (NE 1107b4-8)). It can be seen by this description that temperance—as well as other virtues—occur as part of a triad. The temperate agent (at the centre of the triad) will take pleasure in avoiding over-indulgence (one extreme of the triad), but will also avoid abstemiousness (the other extreme of the triad), and will indulge herself to the right degree for her, as would also be determined by a wise person.

This discussion of excellent character brings out the relationship between emotion, reason and action in determining character type: in the excellent character types, the agent's emotion, reason and action are correct for the circumstances and are also

aligned with each other. However, there are other possible configurations of emotion and reason in determining action, and these are indicative of other character types.

2.4.4 Self-controlled and Weak-willed Character Types

Aristotle's approach to explicating and differentiating self-controlled and weak-willed character types is first to "set out what appears true about our subjects" (NE 1145b4) and then to analyse each type using his theory. What appears true about self-control and lack of self-control is:

> "both self-control and endurance are thought to be good things, and objects of praise, lack of self-control and softness bad, and objects of censure; and self-control is thought to go with sticking to one's rational calculations, lack of self-control with departing from them. Again, the un-self-controlled person acts because of his affective state, knowing that what he is doing is a bad thing, while the self-controlled one knows that his appetites are bad but does not follow them because of what reason tells him" (NE 1145b8-14).

So in each of the self-controlled and weak-willed character types there is an internal conflict between the actions dictated by rational desires on one hand and irrational desires on the other. A so-called self-controlled agent (*enkrates*) will reason and choose to behave in the right (e.g. courageous or temperate) way. Her rational calculations, which indicate that she should behave in the right way, are able to overcome her affective states (such as excessive fear) which push her to act in the opposite way to her choice. In this case behaviour is in line with *prohairesis* but not in line with emotion, thus there is an internal conflict in the agent between the rational and the irrational. In contrast, after a process of deliberation a weak-willed agent (*acrates*) knows, like the strong-willed one, which is the right behaviour in a particular situation, but she is unable to overcome her affective states which push her to behave in the wrong way. She thereby behaves in the wrong way despite choosing rightly (so knows that what she does is wrong)—for example she is either foolhardy, charging into the dangerous situation, or cowardly, running from it. The agent's behaviour here is not in alignment with her *prohairesis*[9].

2.4.5 Bad character

Bad agents (*kaka*) have a fifth type of character in Aristotle's taxonomy. They deliberatively choose to behave in the wrong way and their rational calculations are in alignment with their affective states: both dictate wrong behaviour. For example, the

---

[9] I further develop the Aristotelian account of weak-will (*acrasia*) in chapter three.

bad policeman's choice is to run from the mugging he has stumbled upon, despite not being in disproportionate danger, and his emotions, such as pleasure, are aligned with this choice.

2.4.6 Brutish character

For the sake of completeness in this account I now describe the brutish agent (*akolastos*). The *akolastos* has the last of Aristotle's character types, but it is not clear from what he says if *prohairesis* plays a part in the origin of brutish, or extremely deviant, behaviour. The disposition to be brutish is something that can "come to be pleasant [...] because of disablement or through habituation, or again because of natural lack of quality" (NE 1148b17-19). A brutish disposition can result from disease, such as madness (NE 1148b25), or from habituation, as in the case of "pulling out one's hair or chewing on one's nails, or again on charcoal or earth" (NE 1148b27-28). Each of these causes of a brutish disposition could run against the suggestion that brutish behaviour is preceded by a decision following deliberation: deliberation may be significantly compromised in madness and probably does not take place in trichotillomania. However, Aristotle also mentions cases of brutish behaviour resulting from habituation in agents "who are abused from childhood on" (NE 1148b31). As I have said, habituation involves repeated behaviours with their associated emotions, and if deliberation and decision have been involved in habituation, then deliberation and decision could result in future behaviour that is brutish. Furthermore, in these and other cases, Aristotle argues that "it is possible sometimes, to have [these traits] without being overcome by them - I mean, if Phalaris had restrained himself when he felt an appetite for a child to eat[.]" (NE 1149a12-14). In this example Phalaris might have *prohairetically* deliberated and decided not to eat a child. In sum, it is possible that *prohairesis* is linked to character in some brutish behaviour, though it is the agent's extreme actions that identify brutishness.

Urmson helpfully summarises the role of emotion, action and choice (or rational calculation) for excellence of character, self-control, lack of self-control and badness of character.

Table 3: Summary Table for Sub-types of Character (Urmson, 1973, p.226)

	Emotion displaying a mean state	Action displaying a mean state	Choice displaying a mean state
Excellence of character	Yes	Yes	Yes
Self-control	No	Yes	Yes
Lack of self-control	No	No	Yes
Badness of character	No	No	No

2.5 *Prohairesis* as a Judge of Character

It can be seen from the discussion above that emotions, *prohairesis* and action together indicate the character of the agent. Aristotle makes the claim that there is an important sense in which *prohairesis* "indicate[s] the difference between people's characters more than actions do" (NE 1111b6-7). In defending this claim of Aristotle's, Sauvé Meyer gives two examples demonstrating the greater importance of *prohairesis* than actions in judging character:

> *[P]rohairesis* is a better indication of [...] character than [...] actions because the same action can result from very different *prohaireseis* (plural). For example, George might give money to needy Sam in order to gain a reputation for largesse, while Sandra might do so in order to make sure that Sam does not go hungry. Or James might return what he borrowed because he has been told to do so by his parents, whom he wants to please, while John might do so because he thinks it is the right thing to do. While the first agent in each example performs the action that he should, he does not do so "as the virtuous person would" (NE 1105b7-9: cf 1116a11-15). The deficiency is in his *prohairesis,* rather than in his action. (Sauvé Meyer, 2006, p.140).

So in Sauvé Meyer's examples, Sandra and John not only perform the right actions, but they also choose the right actions for the right reasons—respectively to meet Sam's need for food or to honour a debt. Mele also defends Aristotle:

> "Aristotle has one other reason for saying that choice 'discriminates character better than actions do.' While an action, A, may have all the characteristics of a just action, we cannot, from the mere performance of A, infer anything about the agent's character; for A may have been an involuntary action (1144a13ff. e.g.). But to say that a person's action is chosen is to imply that it was done voluntarily. (1111b7-8). And if we know both *that* an action was chosen and *what* it was chosen 'for the sake of', we are in a position to make an inference about the agent's character" (Mele, 1981, p.414).

Mele is perhaps guilty of overstating his claim, since there could be some things that can be inferred from the agent's actions. If, for example, the agent's action is involuntary but she does not take an option, if one is available, to correct it in some way

then we can infer that she is pleased with the result of her (involuntary) action. If the action is voluntary, then *ex hypothesi,* we know both that the source was in her and that she also satisfied certain knowledge conditions. What is missing in an assessment of the agent's action, however, is whether or not it is in line with her reason or emotions, or with both her reason and emotions.

Wiggins objects to Aristotle's claim that choice indicates character better than actions. Having construed "Aristotle's assertion that choice and deliberation are of what is towards the end (*ton pros to telos*) to mean that choice and deliberation are concerned only with means", he suggests that in order to identify good and bad character, choice would have "to be a fairly inclusive notion that relates to different specifications of man's end." In other words, Wiggins is arguing that if we only know about the agent's choice in relation to means to an end, we cannot infer anything about the agent's choice of end, and surely choice of end is important to revealing the agent's character. Wiggins gives the example of the bad man (*akolastos*) who does not make an error of deliberation about means but who has a "misconception[.] of the end" (Wiggins, 1975, p.31). So in this case the agent's choice—about means—is the right one for the specified ends, but it is the (bad) end that is more revealing about his character. In response, we can return to the wild duckling and truffles example and argue, as above, that deliberation and desire are related to each other dynamically: when an agent deliberates about means she is also deliberating about the end, and the end is "something that seems good to the deliberator" (Sauvé Meyer, 2006, p.139). Sherman's interpretation of Aristotle is that deliberation about what contributes to ends "include[s] deliberation both about the constituents and specifications of an end and about the means to an antecedently fixed end" (Sherman, 1989, p.71). *Prohairesis* can, therefore, indicate character in cases such as that of the *akolastos*.

Interpretation of the action of a weak-willed agent highlights difficulties that may arise when *prohairesis* alone is used to indicate character, because "the flaw will not even show up in the *prohairesis*, for the weak-willed agent is one who acts contrary to his *prohairesis*" (Sauvé Meyer, 2006, p.140). It is the weak-willed agent's action and affective states that miss the mean. Contra Sauvé Meyer, in this case, *prohairesis*, albeit in combination with the agent's emotions and actions, does help to indicate character. This is because unless we know the *acratic* agent's *prohairesis* and emotions she could seem to have the same character as the *akolastos* who acts in line with *prohairesis*.

In this section I have argued, after Aristotle and Sherman, that *prohairesis* is an agent's voluntary and deliberative decision about how to act to achieve, at any time in the future, an end that she desires. Following Aristotle's account, I argue that an understanding of the agent's *prohairesis* gives a better indication of character than mere actions. However, it is necessary to utilise all of the concepts of *prohairesis*, emotions and action in order to demarcate Aristotle's six different character types.

2.6 Conclusion for Aristotle's Account of Character

I have argued that in Aristotle's account, character is a disposition of the soul that may be developed in humans by habituation arising from a natural tendency to experience feelings and emotions. Importantly, character is a voluntary disposition to choose how to act and is also shaped by the choices made by an agent. However, agents are capable of acting as a result of both rational choice and rational and irrational emotion, and in Aristotle's account different configurations of choice, emotion and action contribute to the identification of six different character types. In two of these character types Aristotle makes further stipulations about their features: excellent character types must show intermediacy and also be consistent with the wise person's choices, emotions and actions. Aristotle's account dates from before the common era, so I now defend it against selected, more modern accounts from Kant, Kupperman (1991), Goldie (2004), and Williams (1973). I have chosen to defend Aristotle's account of character against Kant's in particular because they disagree about a fundamental role in character for the emotions, and a fundamental role in character for the emotions is key to my account of weakness of character.

2.7 Kant on Character.

2.7.1 Introduction to Kant on Character.

Kant's ethics has been extensively criticised, and a prominent theme in this criticism is that he does not make a proper place for the emotions in his account of character. An appropriate role for the emotions in an account of character is one that acknowledges, as Aristotle does, their fundamental role in the moral life of agents. So an account of character, like Kant's, that renders the emotions secondary in human motivation is one that denies them this fundamental role. My purpose in discussing Kant is to bring out more clearly the plausibility of Aristotle's account of the role of rationality with desire

and emotion. So in this section I will argue that, despite recognising a role for the emotions in his account of character, the role afforded by Kant ultimately fails to give them their proper place. It fails because his account of character has two component parts—the "intelligible" (*denkungsart* or *noumenal*) and the "sensible" (*sinnesart* or *phenomenal*)—and emotion's role originates in the sensible character, which is secondary to the dominant intelligible character; Kant stipulates that the virtuous agent should be guided only by reason *via* the dominant intelligible part of character.

2.7.2 Intelligible Character

Kant's account of character, as I have said, is divided into two component parts—the intelligible and the sensible. The intelligible part of character functions through the agent's reason. For Kant, reason (or "the will") is "good without limitation" (1998, p.7). So, because the intelligible part of character instantiates reason, Kant says it is *virtuous* character. Intelligible character arises in the agent in "an instantaneous revolution in the will. It is an immediate and full commitment to the moral law, an act which is open to all of us, at any time in our lives, regardless of previous acts, influences or circumstances" (Athanassoulis, 2005, p.114). Kant therefore suggests that we should suddenly become motivated by duty to obey the moral law. Furthermore, intelligible character is very stable: "[m]orally speaking, character is the *steadfast* commitment to virtue that is realised through a *resolute* conduct of thought (*denkungsart*) that is morally good in its form and that, in exercise, entails both causal and reflective elements" (Munzel, 1999, p.2, my italics). Kant himself is explicit about stability of intelligible character in the *Anthropology:* the man of moral character is "[t]he man of principles, from whom we know for sure what to expect, not from his instinct, for example, but from his will, his character" (Kant, quoted in Athanassoulis, (2005, p.127)). Kant's idea of stability of character, or what to expect for the agent, thereby differs from Aristotle's idea, which is that stability is a tendency to behave in the same way in similar situations. It can be seen that Kant also differs from Aristotle in the dominant role he gives to reason in virtuous character.

2.7.3 Sensible Character

In contrast with Kant's "instantaneous revolution" in intelligible character, Kant claims that a gradual change of character in the "sensible" side cannot result in true virtue because "true moral virtue cannot be achieved merely by a change of habits and reform of conduct. This is because any such gradual change, without the underlying revolution

in orientation, is bound to conform to the principle of self-love and not the principle of duty" (Athanassoulis, 2005, p.135). According to Kant, being motivated by "self-love" would tend to lead the agent away from virtuous conduct. So it appears in Kant's account that habits shape the affective, "sensible" side of a person, and this is both subsidiary in moral action and likely to be led astray: "virtue is not to be defined and valued merely as an aptitude and ... a long-standing habit of morally good actions acquired by practice. For unless this aptitude results from considered, firm and continually purified principles, then, like any other mechanism of technically practical reason, it is neither armed for all situations nor adequately secured against the changes that new temptations could bring about" (Kant MS 383-384 quoted in Athanassoulis (2005), p.116). So in Kant's account, the sensible, affective side of character cannot lead to virtue because it is not sensitive to duty and lacks the stability to resist being led astray.

2.7.4 Emotion in Kant's Account of Character

In order to shed more light on the role of emotion in Kant's account of character it is helpful to consider an example from Schiller. Schiller's epigram is an oft-quoted early criticism of the role of the emotions in Kant's ethics, and may be the earliest (Beiser, 2007, p.237).

> "The Scruple of Conscience.
>
> Gladly I serve my friends, but alas I do it with pleasure.
>
> Hence I am plagued with doubt that I am not virtuous.
>
> The Verdict.
>
> For that there is no other advice: you must try to despise them,
>
> And then do with aversion what your duty commands" (Schiller, quoted in Beiser (2007) p.237).

Despite Beiser's claim that Schiller's intention in the epigram is merely to spoof "one common misunderstanding" of Kant's doctrine, the epigram does draw attention to a key problem with the role for the emotions in Kantian ethics. It is best to examine this key problem after quoting Kant himself.

> "To be beneficent where one can is one's duty, and besides there are many souls so attuned to compassion that, even without another motivating ground of vanity, or self-

> interest, they find inner gratification in spreading joy around them, and can relish the contentment of others, in so far as it is their work. But I assert that in such a case an action of this kind—however much it confirms with duty, however amiable it may be—still has no true moral worth, but stands on the same footing as other inclinations, e.g. the inclination to honour, which if it fortunately lights upon what is in fact in the general interest and in conformity with duty, and hence honourable, deserves praise and encouragement, but not high esteem; for the maxim lacks moral content, namely to do such actions not from inclination, but *from duty*. Suppose, then, that the mind of this friend of humanity were beclouded by his own grief, which extinguishes all compassion for the fate of others; that he still had the means to benefit others in need, but the need of others did not touch him because he is sufficiently occupied with his own; and that now, as inclination no longer stimulates him to it, he were yet to tear himself out of this deadly insensibility, and to do the action without any inclination, solely from duty; not until then does it have its genuine moral worth. Still further: if nature had as such placed little sympathy in the heart of this or that man; if (otherwise honest) he were by temperament cold and indifferent to the suffering of others, perhaps because he himself is equipped with the peculiar gift of patience and enduring strength towards his own, and presupposes, or even requires, the same in every other; if nature had not actually formed such a man (who would truly not be its worst product) to be a friend of humanity, would not he still find within himself a source from which to give himself a far higher worth than that of a good-natured temperament may be? Certainly! It is just there that the worth of character commences, which is moral beyond all comparison the highest, namely that he be beneficent, not from inclination, but from duty" (Kant, 1998, p.13,14).

If we apply the reasoning in this quote from the *Groundwork* to Schiller's epigram we can see that the agent may be correct to doubt that he is virtuous if he feels pleasure in serving his friends. Kant (1996) makes this clear in the *Metaphysics of Morals* (6:292-393) when he asserts "a human being cannot see into the depths of his own heart so as to be quite certain, in even a single action, of the purity of his moral intention and the sincerity of his disposition, even when he has no doubt about the legality of his action." Kant is saying that for an agent to be virtuous it is necessary that he acts from duty, and if he feels pleasure in a dutiful act this will not allow him to identify if he is acting from inclination or from duty. The way for the agent to make it clear to himself whether or not he is acting virtuously is to follow Schiller's "verdict" and act from duty to his friends but against his (modified) despising inclinations. However, knowing you are being virtuous is not a necessary feature for Kant of being virtuous, so trying to "despise" your friends is not necessary for virtuous acts of kindness towards them. Despite this, as outlined in Schiller's epigram, there is a significant concern about Kantian ethics: nobody can know that they are acting virtuously unless they derive no pleasure from the act that follows the categorical imperative. This is in contrast with Aristotle whose virtuous agent takes pleasure in acting virtuously.

Baron suggests that Kant, in the long passage quoted above, should have made a contrast between a person who has the right inclination and possesses duty and someone

who has the same (correct) inclination but lacks duty, rather than between a person who has duty only and another who only has the correct inclination. She does this in order to ask if there is "something wanting in a person who lacks a sense of duty" (Baron, 1997). However, Baron's suggested comparison is still open to the criticism that a virtuous agent can never know that he is acting virtuously. This is a significant problem for Kant because our emotions are important to how we appraise many aspects of our circumstances and ourselves, and to remove this as a necessary aspect of our moral lives impoverishes us.

2.7.5 Emotion in the Sensible Character

People attempting to defend Kant against the charge that emotions can play no part in virtuous acts often turn for help to the second of the two parts of character introduced above—the sensible character. The sensible character, in contrast with the intelligible character, is "subject to certain natural tendencies and temperament" (Athanassoulis, 2005, p.128). It is in this part of character, therefore, that the agent may experience emotions that are aligned with reason. Munzel, paraphrasing Kant, describes these emotions as an '"aesthetic quality' of virtue, or its temperament[, as] 'spirited and cheerful.' While a 'slavish attunement of mind' is ipso facto one 'harbouring hatred for the law'; again, 'the cheerful heart in the observance of one's duty .... is a sign of the genuineness of the virtuous disposition (*gesinnung)*.'" (Munzel, 1999, p.305). Baxley, too, cites passages in the Metaphysics of Morals where Kant calls for us to cultivate our emotions if we are to act morally: "general feelings of love and respect, as well as more specific person-directed feelings, like sympathy and gratitude, are allies of duty in the sense of facilitating our ability to carry out our various duties of virtue" (Baxley, 2003, pp.577-578). This, however, takes us back to Schiller, and the claim that an agent whose emotions are in line with reason is unable to be sure that her motivation is virtuous in a Kantian sense.

2.7.6 Comparison Between Kant and Aristotle on the Role of the Emotions

Despite there being a place in Kant for the emotions in facilitating dutiful acts, this role differs from the Aristotelian one where the right emotions and right reason align themselves as partners in virtuous acts. In the sensible character "empirical impulses cannot act directly on the will causing action. They can serve as incentives to action, but they can only determine the will insofar as they are taken up as a maxim." The Kantian agent, therefore, has always to be able to stand outside herself and judge her emotions

as “possible grounds for action” (Athanassoulis, 2005, p.128). Emotions have a secondary role in Kant’s ethics, and the agent who has “achieved pure practical reason […] must constantly be on guard against heteronomy and empirical inclinations” (Louden, 1986, p.481). It is the intelligible character that has prime importance: to be virtuous, a revolution in the agent’s will has to occur first and this then has a cultivating effect on the sensible character.

2.7.7 Conclusion for Kant on Character

In sum, against his critics, Kant does have an account of character which includes a role for the emotions. However, this role, being a secondary one, differs from the fundamental role for emotions in Aristotle. Kant permits a role for emotion in guiding virtuous behaviour, but this may leave the agent in doubt about her (Kantian) virtue. Kant also has an account of character which includes commonly accepted features such as virtue, habituation and stability. However, each of these features has a distinctly Kantian interpretation which deviates from Aristotle’s account: virtue in Kant’s account is in the intelligible part of character, habituation is habit forming of emotions that may go astray, and stability implies determinate behaviours rather than tendencies to behave consistently. Ultimately, however, because Kant’s account of character is divided into two parts with the non-emotional intelligible part being dominant over the emotional sensible part, human emotional life is not a necessary part for him of virtuous conduct. Human emotions, despite their range, depth and value in our lives, are side-lined in Kantian ethics.

2.8 Selected Contemporary Accounts of Character

In this section I analyse aspects of selected contemporary accounts of character to identify potential differences from, and advances on, Aristotle’s account, as outlined earlier in this chapter. If contemporary accounts of character have either deviated from Aristotle or made significant advances, then these would need to be included in a character-based explanation of harms to an agent who is offered an additional option. The modern accounts of character that I analyse were perhaps developed in response to Anscombe who argued, against consequentialist and deontological theories, that if a defensible position in ethics is to be developed then a plausible account of character is needed, since this will explain “how an unjust man is a bad man, or an unjust action a bad one[.]” (Anscombe, 1958, p.5)

Three modern accounts of character are provided by Goldie (2004), Kupperman (1991) and Williams (1973). I have selected these on the basis of their scope (Goldie and Kupperman) and the influence of the writer (Williams). Goldie and Kupperman's accounts of character are more extensive than Williams', and are motivated by the general aim of producing a modern account; Williams criticises both the lack of a place for character in consequentialist and deontological theories of ethics and the concept of a narrative as applied to character. With these differences in scope and purpose in mind, in this section I make the following two claims. First, both Goldie and Kupperman's accounts provide an initially plausible set of three defensible features of character. Character is, first, based on varying constitutive psychological features of the agent—features that the agent is born with; second, it is dispositions that form over time with repeated performances and evaluations of relevant acts; and third, it is dispositions which can be understood in the context of a life "narrative". However, I will argue that only the first two of these three features are defensible. My second claim in this section is that the notion of projects and commitments (Williams, B., 1973)—a fourth feature of character—has a place in an account of character. Following a brief analysis of the four features, I identify for each of them, first, differences from Aristotle's account of character, and second, potential advances on Aristotle's account. A place for constitutive psychological features in an account of character is defensible, and Aristotle has a rudimentary account of this. In contrast, Aristotle has a rich explanation of the role of repeated performances and evaluations of relevant acts in the development of character. Last, applying the notion of narrative to character is not so easily defended, especially from within a life, and Aristotle did not utilise this concept in his account.

2.8.1 Constitutive Psychological Features of the Agent

Goldie (2004, p.31) posits the existence of constitutive psychological features of an agent by referring to a cruel quote from one Labour politician, Dennis Healey, about David Owen, a former Labour politician. Goldie's idea is that a negative trait present at Owen's birth is fixed and that, as a result of the moral weight of this trait, and throughout Owen's life, three other positive traits are undermined. However, Goldie does not shed any light either on the possible origins or on the possible range of such traits. In response to Goldie, there is empirical evidence that people have tendencies to express certain psychological traits: in one theory these traits are Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism (OCEAN) (John and Srivastava, 2010). One objection to the OCEAN model of behavioural tendencies at

birth is that the tendencies may not have moral significance, may be hard to cultivate, and are "too rudimentary to vindicate virtue ethics" (Prinz, 2009). Any such tendencies, however, are likely to provide a substrate for, and may also influence the development of, character; an introvert, for example, may be less likely to habituate emotions and actions in the company of other people. So what has been established here so far is that there is some plausibility to the notion of constitutive psychological features of the agent.

2.8.1.1 Kupperman on Constitutive Psychological Features

The opposite view, that people are not born with constitutive psychological features, is the idea of the *tabula rasa*, or "blank slate" (Locke 1979, p.26). Arguing against the *tabula rasa*, Kupperman develops an analogy between character and a tablet with "lines engraved in a surface", and claims that the idea of a completely blank tablet is an uncomfortable one because this would make one mind "as like another mind as two blank tablets in the tablet storeroom". Furthermore, he adds, "new-born infants are not all alike, everyone has some degree of character even as a baby" (Kupperman, 1991, pp.3,4). Kupperman's claim about constitutive features has plausible elements, since it is an almost universal human experience that no two children, even if they are identical twins, are psychologically alike. For example, we observe that newly born babies' tolerance of stimulation varies, before it has been possible for them to acquire different traits through post-natal environmental influences. However, as a foetus undoubtedly has experiences *in utero*, it is difficult to discount pre-birth environmental influences on any constitutive psychological features a baby might be observed to posses at birth, or to apply the idea of a *tabula rasa.*

2.8.1.2 Aristotle on Constitutive Psychological Features

The first feature of character that I have identified in Kupperman's and Goldie's accounts is that character is based on varying constitutive psychological features of the agent—features that the agent is born with. One aspect of this feature of character is mentioned by Aristotle: he discusses the idea of "natural virtue", which is a type of character present in an agent before any habituation and training has taken place. The state of natural virtue can be contrasted with full virtue which, rather than by chance (or from the gods), Aristotle argues should only come about by habituation and training.

> "[E]ven if happiness is not sent by the gods but comes through excellence and some process of learning or training, it is one of the most godlike things; for the prize and

> fulfilment of excellence appears to be the highest degree good, and to be something godlike and blessed. It will also be something available to many; for it will be possible for it to belong, through some kind of learning and practice, to anyone not handicapped in relation to excellence. And if it is better like this than that we should be happy through chance, it is reasonable to suppose that it is like this, […] To hand over the greatest and finest of things to chance would be too much out of tune" (NE 1099b15-25).

Note that Aristotle also talks here about people "who are handicapped in relation to excellence". He does not clarify in this passage how the person is handicapped, but it is possible that the handicap is a constitutive part of the agent, and present from birth. In a discussion of brutishness Aristotle says this disposition "occur[s] naturally" (NE 1148b30), suggesting that people may be born with the relevant trait. In NE I.9 Aristotle talks about children who "are said to be happy [and] are being called blessed because of their prospects". However, in this passage it is not clear if the child's prospects are good because of her circumstances in life or through her psychological endowment at birth. So this first feature of character in the modern accounts is mentioned by Aristotle, but it has the potential to be both challenged and elaborated in the light of modern knowledge, which supports the notion of constitutive features of individuals. Aristotle would reject Kupperman's claim that babies may have "some degree of character" since, in his account, character can only be formed through habituation.

2.8.2 Dispositions that Form Over Time: Goldie

The second defensible feature of character in the modern accounts is that it is dispositions that form over time with repeated performances and evaluations of relevant acts. Goldie's account of this feature is more comprehensive than Kupperman's, and is based on his claim that the dispositions that form over time with habituation are *reason responsive*. By this he means "that a character trait involves a disposition reliably to respond to certain kinds of reasons—unlike a mere action tendency, behavioural habit or temperament, like being charming, or being fidgety or being gloomy" (Goldie, 2004, p.13). To illustrate this, Goldie gives an example of Susan who, when Miranda drops a book, responds to a "helpful" reason to pick it up. If the act is done out of kindness then Susan "must have a relatively enduring disposition *reliably* to have kind motives and to act in a kind way, so that the appropriate "if-then" conditional can be applied to her: roughly, *if* Susan is in a situation where kindness is appropriate, *then* she will reliably have thoughts and feelings that are characteristic of kindness, and thus will reliably act as a kind person should" (Goldie, 2004, p.15). Goldie grounds this feature of character in an agent's origins "as social animals, [born] into a cultural world of value and

disvalue—a world where certain things *matter*, as harmful, dangerous, comforting, warming and so on. If we have been brought up in the right way, we will be disposed reliably to recognise these values and disvalues and to respond as we should" (Goldie, 2004, p.47). So Goldie is making the plausible suggestion that recognising values is a source for reasons to act, and that these reason-based actions become more reliable with repetition over time. He also acknowledges the importance of being brought up in the right way as a means of guiding and reinforcing values as reasons to act. However Goldie's account of habituation does not mention Aristotle's notion of habituation of emotions.

2.8.2.1 Kupperman on Dispositions that Form over Time

Kupperman also has a place in his account of character for value in the development of dispositions over time; he assigns three stages for this process.

> "The child's creation of the outlines of a character against the background of temperamental and other constraints, the fine-tuning and filling in of details that takes place in late adolescence and early childhood, and later fine-tuning along perhaps with attempts at revision. The strongest examples one encounters of what sounds like conscious control concern fine-tuning and filling in of details, especially in late adolescence and early adulthood. An important part of this process is articulation of values and ideals" (Kupperman, 1991, pp.55,56).

His account of development of character, therefore, is based on values and ideals held by the agent, but does not, like Goldie, include mention of "reasons" which are generated from these values. He acknowledges a role for habit in the formation of character but, following Hume, is sceptical that an agent can have significant control over the development of her character, except in the fine-tuning phase. He says that his conclusion "that people have little control over their characters is compatible with cases in which an individual has a fair amount of control over some aspects of character. Cases of very little control, however, represent the norm" (Kupperman, 1991, p.57).

A problem for Kupperman, in the light of Goldie's account, is how, if values and ideals are important to the agent, these can only have a limited role in the development of character. Kupperman's response to this objection is that an agent *can* have control over the development of her character based on her values and ideals, but this is a matter of degree. For example: "I can accept or reject the dominant values of my social class. If I lose my job and have to work in a factory assemble line, I have some control over what to think as I work and how I organise my free time." However, Kupperman continues, "[t]o say that a person has control over any given one of a number of factors that

contribute to a change of character is not […] to say that a person has control over the whole process" (Kupperman, 1991, p.57).

The notion of character as dispositions that form over time with repeated performances and evaluations of relevant acts is fundamental in Aristotle's account. In one example "we become just by doing just things, moderate by doing moderate things, and courageous by doing courageous things" (NE 1103b1-2). Furthermore, Aristotle has a rich account of voluntary choice and responsibility for character, also explained above. So modern accounts of character can provide further support for Aristotle's view.

2.8.3 Life as a Narrative

The third defensible property that I have identified in modern accounts of character is that dispositions can be interpreted in the context of a life "narrative". The term narrative, in the context of dispositions of an agent to act and feel, means that acts and feelings are related to one another within her life because there are plausible causal explanations for their development and relationships to one another. E. M. Forster, quoted by Goldie, uses a royal example to illuminate the idea of narrative as revealing causal connections. "The King died and then the Queen died" is not a narrative because the relationship between the two events is not clear. However, if it can be said that ""[t]he king died and then the queen died of grief" [this] does reveal the causal connection between the two events" (E M Forster quoted in Goldie (2004 , p.114)). Furthermore, the idea of causal connections between events can also potentially be applied throughout an agent's life. Goldie claims that the idea of a life narrative can play a formative role for an agent if she can "take an evaluative perspective on [her] past or future selves" (Goldie, 2004, p.117). An agent may evaluate and regret past actions, and resolve to behave differently in similar situations in the future. In this way a narrative can be identified which consists in causal connections between the agent's evaluation of her past actions and attempts by her to modify her dispositions to act[10].

Without ever using the term "narrative", Kupperman also claims that character should be understood in part as a set of related features within a whole life. He says "[a]n ethics of character also must take account of the ways in which projects and decisions are

[10] The concept of life as a narrative has been applied in therapeutic settings to help people make positive changes in their lives, for example by adapting to changing circumstances or by encouraging changes in negative behaviours, such as addictions or so-called personality disorders. An example of the latter is described by the subject in "Girl Interrupted" (Kaysen, 1995).

integrated through time" (Kupperman, 1991, p.v). He adds that "[c]haracter has a great deal to do with how we are prepared to maintain, modify or abandon a structure of goals or commitments", and that, as well as "bridg[ing] thought and action" "[c]oncerns and commitments provide temporal bridges" (Kupperman, 1991, pp.13,17). Kupperman is describing character using the same sense of narrative as described by Goldie. His "bridges" provide the explanations, and are built throughout a life between "concerns and commitments", which are in turn based on values held by the agent. Kupperman could have used the narrative metaphor in this part of his account, but it was not necessary for him to do so.

2.8.3.1 Williams on Life as a Narrative

It is helpful at this point to draw on Williams' account of narrative in "Life as Narrative" (2009). Williams is sceptical that it is possible from the agent's perspective for her to live her life according to the structure of a narrative, as suggested explicitly by Goldie and implicitly by Kupperman. His claim (with my italics) is that "in understanding people's lives—above all, *other people's lives*—narratives that give them a certain direction or meaning are very important" (Williams, B, 2009, p.312). However, for the agent in the context of her own life, it is not so easy for her to understand her life and to plan to modify her dispositions on the basis of evaluations of past actions. The potential for difficulties with forward planning in this way was described by Kierkegaard, who is quoted by Williams (2009):

> "It is perfectly true, as philosophers say, that life must be understood backwards. But they forget the other proposition, that it must be lived forwards. And if one thinks over that proposition, it becomes more and more evident that life can never really be understood in time simply because at no particular moment can I find the necessary resting point from which to understand it—backwards" (Kierkegaard 1843 in Williams, B, 2009).

Kierkegaard's claim is difficult to interpret, but it could be that, because the agent is constantly moving forwards in her own life, there is no stationary point from which to make evaluations. Or perhaps he is simply claiming that it is not possible to "understand" a life, or reason about life events that have not yet taken place. However, Goldie's example (above) of the agent who regrets her past actions and resolves to behave differently in future may demonstrate that the clock can be stopped to permit reflection on a particular life event or set of events. Furthermore, it is not implausible that future similar circumstances may present themselves to the agent making it possible for the agent to understand her life forwards. Goldie's example is a simple one, and, whilst it

can be applied to specific incidents, it is harder to aggregate such examples in an agent's life and extend them over larger time periods. This difficulty is the source of Williams' conclusion that life is naturally disorderly, which consequently undermines "the supposed coherence and unity that narration can give to people's lives" (Williams, B, 2009, pp.312-313). So, it seems impossible to conceive of an agent who is capable of narrating all aspects of her life into the future, though perhaps not impossible for her to apply the concept of a narrative in planning isolated aspects of her life.

Despite being aware of the concept of plot (*muthos*), Aristotle does not use the notion of *muthos* in his writings about character.

> "Plots are either simple or complex… The action, proceeding in the way defined, as one continuous whole, I call simple, when the change in the hero's fortunes takes place without Peripety [reversal] or Discovery; and complex when it involves one or the other, or both. These should each of them arise out of the structure of the plot itself so as to be the consequence, necessary or probable, of the antecedents. There is a great difference between a thing happening *propter hoc* and *post hoc"* (Aristotle Poetics quoted in Velleman, 2003, p.2).

This quote shows that since *muthos* may include chains of events which arise from each other, it incorporates the same ideas as a narrative.

In sum, the idea of a narrative structure in the context of character may not be an advance on Aristotle. Narratives about lives are one way of thinking about explanatory links between parts of that life, but they do not add anything to the explanations themselves. Furthermore, narrative explanations may not be useful to the agent in the course of living her life as a whole, but are more easily applied to a person's life from outside. Aristotle assigns to practical reason the role of structuring an agent's life as a whole: "practical wisdom is a capacity which presupposes virtuous character and an attachment to the right sort of ends. As a rational power it includes an ability to form *prohairesis*—reasoned choices about particular actions as they fit *within some overall system*" (Sherman, 1985, p.105, my italics).

2.8.4 Projects and Commitments in the Context of Character

I have said that Williams does not set out to give a complete account of character, but instead makes claims relevant to character, such as those in "Life as Narrative" mentioned above. In "Persons, Character and Morality", which is primarily an argument against Kantian and Utilitarian ethics on the grounds that both theories leave insufficient space for the individual, Williams writes: "[i]f Kantianism abstracts in

moral thought from the identity of persons, Utilitarianism strikingly abstracts from their separateness" (Williams, B., 1973, p.3). His approach to problems with abstraction from the identity of individual persons and abstraction from separateness of persons "involve[s] the idea that an individual person has a set of desires, concerns or, as I shall often call them, projects, which help to constitute a character" (Williams, B., 1973, p.5). It is not necessary for my purposes here to go into the detail of Williams' arguments against Kantianism and Utilitarianism, but I do need to say more about the claim that projects help to constitute character.

2.8.4.1 Kupperman on the Relevance to Character of Projects and Commitments

Kupperman objects to Williams' claim that projects help to constitute character on the basis that projects and commitments valued by the agent may change and her character remain unchanged. In support of his claim he gives examples, first, of a dogmatic and unreflective communist who becomes a dogmatic and unreflective extreme right wing conservative, and second, of a woman who re-marries.

> If Bludgeon, who has been a dogmatic and unreflective Communist, suddenly becomes a dogmatic and unreflective extreme right-wing conservative, we might well say that Bludgeon's character has not changed at all. If the next time we see O'Reilly she is married to a different man, this is not normally grounds for attributing a change of character. On the other hand, if Bludgeon or O'Reilly has a new policy with regard to taking other people's money, this counts as change of an element of character. (Kupperman 1991, p.11).

Kupperman concedes that projects "certainly reveal character; nevertheless, our ordinary conception of character is independent of *what* someone's projects and commitments (apart from specifically moral commitments) happen to be" (Kupperman 1991, p.10). Kupperman seeks here to separate certain projects from the sphere of an agent's character on the basis that they do not fall under the moral realm: he sees an agent's political persuasion and choice of partner as being non-moral and thereby not part of character. However, in the political case, communism and right wing politics consist in sets of values which, by their nature are moral. So in this case, *contra* Kupperman, we might have grounds for suspecting that something dramatic is happening to Bludgeon's character. However, in the marriage case, O'Reilly's behaviour lends support to a view that she values marriage, whether with one man or the next. We are right, then, to agree with Kupperman that O'Reilly's character has not changed, unless of course there is a key aspect of the first relationship that is clearly

moral—strict monogamy, perhaps—and which differs from a moral aspect of the second relationship—e.g. an open relationship.

2.8.4.2 Aristotle on Projects and Commitments

I have argued above, after Sherman, first, that where Aristotle talks about projects and commitments in the context of character, the examples he uses are technical ones. Second, it can be seen however, also after Sherman, that *prohairesis* may apply to non-technical projects and commitments which thereby reflect the nature of the agent's character. So Aristotle's account of character does have a place for projects and commitments.

2.9 Summary on Contemporary Accounts of Character

In this section, I have identified three modern accounts of character and abstracted four potentially defensible features from them. These features have been analysed and then compared with Aristotle's account of character. Three of the four features—constitutive features, dispositions that form over time, and projects and commitments—are defensible and appear both in Aristotle and in the modern accounts. Aristotle's account of constitutive features, however, is under-developed. The fourth feature—narrative—appears in Aristotle's Poetics, but not in his Ethics. Furthermore, the notion of a narrative does not appear to provide an advance on his account; Williams is justifiably sceptical about an agent's ability to prospectively integrate a complex array of projects and commitments into a life as a narrative. The main differences between Kupperman's account and that of Goldie are his pessimism about the scope for an agent to make significant developments in her character, and his implausible separation of certain aspects of character from the moral realm, such as an agent's projects. Though Williams' account of character is incomplete, he also identifies plausible features of it, which are "projects and commitments" valued by the agent. In sum, the three modern accounts of character that I identified do not challenge Aristotle's account.

I have analysed a primarily Aristotelian account of character, but the notion of character itself has faced objections, and key amongst these are objections arising in situationism; situationism is a thesis that the determinants of behaviour are situational and do not arise in dispositions of agents.

2.10 Situationism

The challenge posed by situationism to character based theories is that the origins of behaviour are traceable to the situations in which the agent is behaving rather than to the agent's character. In other words, situationism claims that our dispositions are very weak predictors, when compared to situations, of how we behave. Contrary to this, dispositionism states that an agent's dispositions, including her emotions and choices, can be more predictive of her behaviour. I defend a claim that dispositionism has the resources to successfully meet the challenge from situationism.

Since dispositionism incorporates the commonly held idea that moral character is a strong determinant of our behaviour, situationists claim that we have reason either to dispense with the idea of moral character (Harman, G., 1999) or at least to radically revise it (Doris, 2002). The argument between situationism and dispositionism seeks to base moral theory about human behaviour in an account that is "psychologically possible" (Flanagan, 1991, p.26). The situationists claim that empirical research shows character dispositions are not psychologically possible—i.e. humans have been shown not to possess dispositions.

2.10.1 Empirical Studies Purporting to Support Situationism

I describe four of the principal empirical experiments on which situationism is based in order to show how dispositionism can meet the challenge they pose. First, is the "Honesty" study of children (Hartshorne and May, 1928). The "situation" in this experiment consisted of the circumstances in which a child was placed, which, for example, was finding money in an empty classroom or marking her own exam paper. The outcome of the study was that subjects did not show consistency in their behaviour between the different settings. So the situationist claim here is that the children in the experiment did not show character dispositions—e.g. to be honest or dishonest—that manifested themselves consistently in different settings.

The second example is the "Dime Finding" experiment (Isen and Levin, 1972). The situation for the subject here involved being faced with someone who had dropped some papers, after the subject had either found a dime or not found a dime. The outcome of the experiment was that people were more likely to help the person who had dropped some papers if they had found a dime before encountering her or him. So the disposition being tested here is that of being helpful. Isen and Levin claim that the

results refute the significance of such a disposition and that the situation—namely finding a dime—is the determinant of the subject's actions.

The third experiment—"Samaritans" (Darley and Batson, 1973)—is similar in nature to the second. The situation being investigated here is one in which either the agent is in a hurry to give a talk about a morally laden subject, or she is not in a hurry, and then she encounters a person in distress. The outcome was that being in a hurry made the subject less likely to help. Purported dispositions held by the subjects—such as those that may be related to being religious—apparently had no bearing on the behaviour of the subjects.

Last in my list is the "Obedience" experiment (Milgram, 1963). Subjects in "Obedience" were given instructions by a sham researcher to shock further subjects who did not answer questions correctly in an experimental situation. The claimed outcome for the study was that it was the situation of being given instructions that determined the behaviour of the subjects, rather than the subjects dispositionally behaving non-harmfully, as had been predicted before the experiment.

On the basis of these and other experiments Doris and Harman in particular challenged dispositionism. Harman claims that because the behaviour of the subjects in the situationist experiments does not appear to be predicted by their dispositions as construed by the researchers, "ordinary attributions of character-traits to people [—the fundamental attribution error—] may be deeply misguided, and it may even be the case that there is no such thing as character" (Harman, G., 2000, p.165). Doris's claim is that rather than having no character at all, our characters are fragmented: "we are justified in attributing highly contextualised dispositions or "local" traits" (Doris, 2002, p.64). Doris also claims that minor aspects of situations can have a disproportionate effect on behaviour. It can be argued, however, that character based theories do have the resources to meet the situationist challenge. There are at least five arguments against situationism using character theory; some of these arguments rest on methodological concerns about the four experiments.

2.10.2 Responses to Situationism

In the first of the experiments that I listed, the researchers looked at dispositions in children, who are not normally expected to have developed stable characters. There is some evidence from psychology, mentioned above, that children may possess traits i.e.

characteristic innate psychological features such as the "big five"—OCEAN (John and Srivastava, 2010)—but even if children do have traits of this kind, they are not normally thought of as possessors of more stable dispositions, such as a character. So Hartshorne and May should not have expected their subjects to behave as if they were in possession of a character.

Second, the listed experiments test behaviours in situations that are novel to the subject, and also fail to test character traits over more than one instance of behaviour; a more realistic way of testing character would be in a familiar or repeated situation. The situation in Milgram could be defended as a novel one—people are not normally asked to conduct memory experiments. However, people are familiar with electric shocks and the concept of harm from giving electric shocks. Perhaps the novelty in this situation is the experimental setting in combination with the authority of the researcher giving the instruction to shock. Furthermore, none of the experimental situations were repeated. Many of the subjects in Milgram were traumatised by their experience (Levine, 1988, p.218), and would no doubt have behaved differently if re-recruited to a similar trial.

Third, the experiments attack an unrealistically narrow construal of character: they ignore nuanced accounts of character including different sub-types such as *acrasia* and *enkrateia*. Athanasoulis gives examples of two hypothetical agents, A and B, in the "Samaritans" case, who are respectively vicious and weak willed in response to the situation (Athanassoulis, 2000); this more sophisticated dispositional explanation of the subject's responses is plausible.

Fourth, as Annas (2005), Kamtekar (2004) and Miller (2003) point out, the experiments do not recognise and accommodate the role of practical reason in synthesising related local character traits into global ones. Doris defends an account of highly localised character traits. However, these are so localised, and presumably so numerous, that they cannot be construed as character traits at all; they seem instead to be separate instances of situationally determined behaviour. If character were determining the behaviour in each of the numerous different situations, the agent should be able to recognise relevantly similar salient features of the situations and develop through responding in consistent ways. In this way local character traits can be synthesised into more global ones. (Annas, 2005).

Last, the situationist experiments pit different virtues against one another, which is a more severe challenge to character than pitting virtues against their respective vices (Kristjánsson, 2008). Balancing conflicting virtues is more difficult than balancing a virtue with its respective vice because in the latter case, reasoning about what to do and responding to the relevant emotions will be in the same sphere; in contrast, the relevant parts (emotion and choice) of conflicting virtues come from different spheres. Aristotle's response could be that conflicts between virtues should not be problematic because if an agent possesses one virtue then she possesses them all through having practical reason: "for with the presence of the one quality, practical wisdom, will be given all the virtues" (NE 1145 a1-2). Practical reason, in turn enables the agent to "perceive that, in this particular case, the virtues do not make opposing demands or that one rule outranks another, or has a certain exception clause built into it" (Hursthouse and Pettigrove, 2016). However, given the extreme rarity of virtue (and thereby practical reason), conflicts in the realms of two different virtues will be more problematic, for the reasons given above, than balancing a virtue with its respective vice.

So far in this chapter I have explained Aristotle's account of character and compared it with more modern accounts. I have also defended character-based theories against objections arising in situationism. What has emerged in the above is that a character-based moral theory provides a rich account of human behaviour, including an explanation of its origins, development, responsibility for actions, and a taxonomy of character that, for example, makes space for concepts such as weakness of will. These are all reasons for assigning a central role for character in moral theory, which I now defend in more detail.

2.11 A Central Role in Moral Theory for the Concept of Character.

It is important for my overall thesis that character has a central role in moral theory. Without this central role it is not possible for me to use the concept of character, or more specifically forms of weak character, to explain harms to an agent who is offered an additional option. I claim that character should have a central role in a moral theory. This is because the notion of character provides a rich means to understanding the moral behaviour of agents in at least three different ways: first by explaining the origins and development of behaviour, second, by explaining the agent's responsibility for her behaviour and thus her assessability as blameworthy or praiseworthy, and third, by further illuminating moral behaviour using different character types. I have already

shown how Aristotle's account of character can meet these desiderata. I have also responded to an objection to character based theories from situationism. Louden (1984), however, objects, first, that character-based theories lack the ability to guide action, and second, that character-based theories do not enable an evaluation of actions that are "out of character" (Louden, 1984, p.229). Furthermore, I have not established that an analysis of character enables the evaluation of action in ways that are not possible in moral theories using only ideas of right and wrong. I now respond to these objections in a way that is commensurate with a dissertation that does not have this area as its main focus.

2.11.1 Louden on Action-guidingness

Louden claims that the "skills of moral perception and practical reason are not completely routinisable and so cannot be transferred from agent to agent as any sort of decision procedure 'package deal'". If Louden is correct, then a character based theory cannot be used to guide good behaviour in this way. One of Louden's examples of a character theory—virtue ethics—failing to guide action is of "the doctor/patient principle of confidentiality [which] must always (or not always) be respected" (Louden, 1984, p.230). Louden's objection is that it is not possible, say, for a junior doctor to be merely told using a virtue ethics framework how to behave in a patient confidentiality situation. In brief reply to Louden, in order for the doctor to be guided how to act in confidentiality situations it is necessary for her to habituate her emotions as a product of reasoning and acting under guidance from a wise doctor, in the way that I have described above in the section on Aristotle's account of character. Furthermore, Hursthouse (2006a, p.106) also responds to Louden with an account, following Aristotle, that lists the many different descriptions of types of excellent character—"v[irtue]-rules"—and also types of character that fall short of virtue, arguing that these different descriptions offer greater action-guidingness than accounts that are either rule-based or based on maximising the good. Hence the virtuous doctor will have more to guide her than blanket rules about confidentiality which may not fit the particulars of the case at hand. Importantly for my purposes, if it is possible for a character based theory to guide good behaviour, the same theory may help to explain bad or harmful behaviour. Bad or harmful behaviour of any degree in this case will be behaviour that falls short of good behaviour.

2.11.2 Evaluation of Actions

My second claim in this section is that it is a role of character in a moral theory to allow evaluation of actions by agents that goes further than an evaluation of the act itself using consequentialist or deontological theories. In making such evaluations using either consequentialist or deontological theories, how the act itself measures as a means to increase the overall good, or as a response to a particular rule, informs the evaluation. In evaluating an agent's behaviour with a character-based theory, however, her character as a whole is potentially available for evaluation. Moral evaluations based on character may allow identification of aspects of the agent's character for which she is not responsible, such as certain of her constitutive psychological features (as discussed above), and also permits an evaluation of the agent according to internal psychological tensions or their absence. This is an advance on theories using only the concepts of the right and the good because it offers a deeper evaluation of the agent. Depth of evaluation in this sense is an understanding of the underlying psychological states of the agent prior to acting.

2.11.3 Louden's Second Objection

Louden (1984) also objects that it is necessary to evaluate acts in isolation from agents where the acts are out of character. He says there is a danger that virtue ethics may blind itself to wrongful conduct "simply because it views the Oedipuses of the world as honourable persons and because its focus is on long-term character manifestations rather than discrete acts. To recognise the wrong in Oedipal behaviour, a theory with the conceptual tools enabling one to focus on discrete acts is needed" (Louden, 1984, p.230). In reply to Louden, it can be argued that theories of character can accommodate "acting out of character". As I have said, it is not normal for dispositions of character to be totally stable, and acts that would not have been predicted because they are out of character for the agent should not be totally unexpected. It is also not possible for all new situations facing the agent to be sufficiently similar to situations she has previously encountered for her to behave in ways that are completely predictable. For example, time will always have passed before the new situation is encountered, and this, on its own, may alter the balance of evaluations of reasons used by the agent in her decision-making process. There is no need to abandon character in these situations, and it is a strength of an account of character that it can enable deeper evaluations of responsibility for behaviour, for example by reference to psychological divisions within the agent.

Louden extends his objection about acts that are out of character by talking about "character change". He claims that character can change because character traits that are acquired through habituation may be lost if they do not continue to be practised. Even if we do not lose them completely we may "lose sensitivity". "Once we grant the possibility of such changes in moral character, the need for a more 'character free' way of assessing action becomes evident. A more reliable yardstick is sometimes needed" (Louden, 1984, p.231). Resources in an account of character that explain the development of character can be harnessed in reply to Louden's further objection. For example, Scarre argues that

> "[u]nless virtue is confused with a kind of psychological truncation, the virtuous person is not one in whom the bodily appetites and other natural desires that can lead one astray have conveniently been extinguished. To err is human because to desire is human. So, for instance, the virtuous person may still be subject to certain sexual impulses that, since they can lead to immoral behaviour, will have to be managed. If virtue is not mere insensibility, then the virtuous person still needs to govern her natural passions, even if she has become more accomplished (firm, decisive, swift) through training and practice at doing this than the moral tyro. What she has *not* done is transcend the need for continence, which would mark her as either a god or a zombie" (Scarre, 2013, p.15).

Scarre has described how a "character free" account is not necessary to explain possible problems posed to a moral theory by character change. It may be better, however, contrary to Scarre, to accept that people cannot be fully virtuous, and that most are either *encratic* or *acratic*. Furthermore, Scarre has in fact described an *encratic* agent rather than a virtuous one. However the reply to Louden is still intact since there is no change in character if an agent switches between *acrasia* and *encrateia*, and this switching can explain what for Louden are either acting out of character or character change.

2.11.4 Identifying What is Moral in Ways that are Not Derivable from the Right and the Good

Last in this section, I claim that character has a role in a moral theory in identifying what is moral in ways that are not derivable from concepts of the right and good. Rawls claims that "there are just two types of ethical theory: deontology is an example of one, utilitarianism of the other. The reason why there are just two types is that there are two 'main concepts' or 'basic notions' in ethics, the 'right' and the 'good', and the differentiating structures of the two types are (largely) determined by how they define and connect those two" (Rawls in Hursthouse 2006, p.100). However, Watson (1990) and Hursthouse (2006a) claim that looking at moral theory in terms of the right and the

good misses an important dimension to moral theory offered by character. This is the dimension provided by character theory that offers an understanding of the origins, development and responsibility for action.

Kawall (2009) also defends character based theories which include the concept of virtue against Rawls' objection that virtues are merely derived from concepts of right and good and are, thereby, not in a position of primacy in moral theory. He says that Rawls's objection could be stated in the following terms: "we identify virtuous persons (or their possession of given virtues) through their performance of certain kinds of actions; as such, these right actions are explanatorily basic or primary, and the virtues are derivative and best understood as dispositions to perform these actions" (Kawall, 2009, p.3). One line of response to this objection offered by Kawall draws on an analysis of right actions that might be expected of a person who is romantically in love. A rule based account of love might say that "loving people are simply those who are disposed to perform independently grounded loving actions." However, this does not seem to offer a satisfactory account of love as "a mere disposition to perform particular actions or follow a set of rules" rather than "loving actions [being] the actions that flow out of the rich emotional state of being in love" (Kawall, 2009, p.4).

Kawall offers a "similar reductive consequentialist understanding" of love as being "disposed to perform actions that maximize the well-being of the beloved." This account can be seen to fall foul of the same criticism as the deontological, right based one since it, too, does not capture a "rich emotional state" (Kawall, 2009, p.4). If the analogy between love and good character is accepted, in sum, neither the right nor the good offer a satisfactory grounding for the behaviour of an agent either with excellent character or with character that is not excellent: character can explain moral behaviour in ways that are not derivative from the right and the good. As Watson says, "[b]asic moral facts are facts about the quality of character" (Watson, 1990, p.452).

So character is a rich psychological disposition that affords moral theory the possibility of explicating and understanding responsibility for behaviours of agents that goes beyond moral theories dependent only on ideas of the right and the good. A moral theory based on character can be action guiding, along with consequentialist and deontological theories, but offers assessments of agents with greater depth than those made on the basis of mere acts.

2.12 Conclusion

I have explained Aristotle's account of character in the Nicomachean Ethics (2002) and defended it against both Kant's account—which has only a secondary role for emotion—and the challenge from situationism. Aristotle's account is a psychologically rich one that consists in habituated dispositions to feel, choose and act, and which is capable of explicating the behaviours of agents in empirical (situationist) studies. Three contemporary accounts of character (from Kupperman, Goldie and Williams) have features in common with Aristotle's account, but the only area where there might be any significant advance made on Aristotle is with the notion of constitutive features of the agent that are shaped into character. Last, the notion of character is a significant feature of a moral theory since it confers the ability to assess an agent's acts in relevant ways that are more sophisticated than moral theories based only on notions of right and good.

Now that I have an account of character in place, I am in a position to defend, in chapter three, an account of "weak character". The concept of weak character, as analysed here, is, as I have said, helpful in explaining potential harms to an agent who is presented with additional options, since the relevant types of weak character may make the agent susceptible to such harms. This line of argument will be relevant in chapter six to cases where the additional option is Physician Assisted Suicide.

Chapter 3. Weak Character

3.1 Abstract

In chapter three, I argue that there are three types of character that may render an agent more susceptible to harms from being given an additional option. In particular, I defend accounts of *acrasia* or weakness of will, undue self-deprecation—a vice which is the defect of character associated with the virtue of humility—and undue lack of confidence in one's judgements—another vice. *Acrasia* influences an agent to choose an additional option, once it is made available as an extension of those she can select from, when she knows it is *not* best for her, but is drawn towards it because of her emotions and less rational desires. Undue self-deprecation may influence an agent to choose an option, once it is made available as an extension of those she can select from, when her choice is simply a manifestation of her vice of taking herself to have unduly low worth. Last, undue lack of confidence in her judgements may influence an agent to either find it difficult to justify her choices or to let go of her values too readily, when she is making a choice.

3.2 Introduction

In chapter one, I defended a claim that an agent may potentially be harmed if she is offered an additional option. I argued that an agent is more likely to be harmed in such a choice situation if she has any of three types of weak character, or if the context of choice is configured in a certain way. Furthermore, these two harms are distinct from harms defended by Dworkin, G. (1988) and Velleman (2015). The three types of character I posited were, first, having a weak capacity for judgement and choice, second, having unduly low self-esteem, and third, having undue lack of confidence in one's judgements that flow from one's values.

In order to be able to give a full account in chapter three of these three types of weak character, it was necessary for me to defend in chapter two an account of character itself. Importantly, the Aristotelian account I defended consists of character as a psychological disposition (*hexis*) to act, but where the disposition is not completely stable and, also, fundamental roles for choice (*prohairesis*) and emotion (*pathe*) and desires (*orexis*). The importance in the notion of character of relatively stable dispositional traits related to a virtue, with fundamental roles for choice and emotion, is that these aspects provide

a framework with which I can defend a full account of my three proposed types of weak character. A further aspect of character is that different virtues and their associated emotions and character-types are relevant in different types of situation; so courage, for example, is relevant in situations that are frightening for the agent. So here I claim, first, that weakness in the capacity for judgement and choice is Aristotle's so-called weak-willed (*acratic*) character type. Then, I will argue, second, that unduly low self-esteem or self-deprecation is an ethical vice in the domain of the virtue of humility (Richards, 1988) which is relevant in choice situations concerned with the agent's needs. Finally, I will claim, third, that undue lack of confidence is a vice, and might be construed as an ethical vice.

A further important aspect of character is that an agent's character influences her actions in the ways that I have defended. So if an ethical theory is to be able to make moral assessments of an agent's actions it needs to take account of character. I am concerned in this dissertation with harms to an agent who is presented with an additional option. As I have said, if an agent is presented with an additional option there may be a harmful impact on the agent's ability to make choices, or alternatively the agent may be more likely to make a choice that is harmful for her. In the latter case, the notion of weak character helps to explain the reason that the selection the agent makes is harmful to her: her selection is not the one that she has reasoned is best for her, or second, her selection is simply a vicious act (reflecting either of the two vices identified above), and she would have been better off not having the option to manifest those types of weak character.

3.3 Weak Will (*Acrasia)*

First, I turn to *acrasia*. The defining feature of being *acratic* in Aristotle's account, as I said in chapter two, is that the agent acts in line with her emotions (*pathos*) and irrational desires (henceforth just "irrational emotions") and against her choice (*prohairesis*) (NE VII. 7). So, if an *acratic* agent is offered an additional option this may harm her if her choice is to act a certain way, but this conflicts with her irrational emotion which influences her to act against her choice: there is an internal tension between choice and irrational emotion which is won by the latter. It is necessary for my purposes to explain how an agent may act in line with her irrational emotion and against her choice, and also to give an account of the relationship between *acrasia* and *enkrateia,* since in both these psychological states there is a tension between choice and

irrational emotion. In the former state, as I have said, irrational emotion and not choice influences action, and in the latter, choice influences action by overcoming irrational emotion (strength of will).

Importantly for purposes connected with my argument about PAS, I claim that an agent may be more susceptible to *acrasia* in situations, first, that may generate strong emotions, second, that are novel, and third, that present her with an opportunity to exercise this aspect of her character. First, an agent who may be prone to oscillation between *acrasia* and *enkrateia*[1] may be swayed towards *acrasia* if she finds herself in a situation that generates strong emotions. In this case, strong emotions may be more likely to overcome her choice. Second, in a novel situation an agent's ability to choose may be impaired because she does not have access to full knowledge about the particulars of the situation with which to deliberate and choose; this is one interpretation of Aristotle's account of *acrasia*. In this second case, since the agent's ability to choose is diminished, irrational emotion is more likely to influence action. Furthermore, the second case, in which lack of knowledge may allow irrational emotion to overcome reason, is relevant to the debate between Socrates and Aristotle on the possibility of *acrasia*. As I will show, Aristotle initially denies, *contra* Socrates, that ignorance on the part of the agent is the cause of *acrasia*, but later seems to make a partial concession. Last, if an agent with a tendency towards *acrasia* is offered an option that is aligned with her emotions then this will present her with an opportunity to exercise her *acratic* tendency.

Before I explain the three ways in which *acrasia* may potentially result in harm to an agent who is offered an additional option, it is necessary for me to say more about *acrasia* itself. I will shed some light on the detail of Aristotle's account because it helps to explain how emotion may overcome reason. First, I will outline the argument between Socrates and Aristotle on the possibility of *acrasia* and then, after following Aristotle's line, I will explain his different types of *acrasia*, which arise in two key emotions. I can then use these types to explain how an *acratic* agent may be harmed in certain choice situations.

---

[1] Most people fall under this category: it is accepted that the other character types, such as virtue, are rare.

3.3.1 Aristotle and Socrates on *Acrasia*

Aristotle's point of departure in his account (NE VII. 2) is to re-state Socrates' denial of the possibility of *acrasia*:

> "But one might raise the problem: in what sense does a person have a correct grasp when he behaves uncontrolledly? Well, some deny that it is possible to do so if one has *knowledge*: it would be an astonishing thing if, when knowledge is in us—this was Socrates' thought—something else overpowers it and drags it about like a slave. For Socrates used completely to resist the idea, on the grounds that there was no such thing as behaving uncontrolledly; no one, he would say, acts contrary to what is best while grasping that he is doing so, but only because of ignorance" (NE 1145b21-8).

So Aristotle reports Socrates as arguing that there is no possibility that an agent's irrational emotions may overcome her knowledge of what is best to do (or her choice): an agent acts in line with her irrational emotions only if she is ignorant and thereby unable to deliberate and choose the correct action. Aristotle's initial response to Socrates is that this interpretation is not consistent with what "patently appears to be the case" (NE 1145b28). However, he adds that it is necessary to take into account details about the agent's emotional state and any potential types of ignorance that may be involved. In his discussion of relevant types of ignorance, Aristotle's eventual conclusion appears to be in partial agreement with Socrates—irrational emotion may only overcome reason if there is ignorance, or deficiency of knowledge, of a sort.

Aristotle defends his account of ignorance in the case of *acrasia* by distinguishing different aspects of knowledge, which I paraphrase from Price's (in Blackwell Guide) list of quotes from Aristotle (2006, p.236). First, people who have knowledge can either consider it and then use it, or not consider it and not use it (NE 1146b31-5). Second, there are two types of premise—universal and particular—so a person may act against knowledge if he uses the universal premise but not the particular one (NE 1146b35-1147a4). Third, the universal premise can apply either to the person, "e.g. 'dry food benefits every man' and 'I am a man'", or to the object, "e.g. such-and-such food is dry", and either type of knowledge may or may not be actualized (NE 1147a4-10). Last, there are different ways in which a person may have knowledge. He may have knowledge and be able to recite it in a similar way to a person who is under the influence of sentiments such as anger and sexual appetite, who has knowledge but does not use it, and does not have knowledge as "second nature" (NE 1147a10-24). Price's list sheds light on how Aristotle subdivides knowledge in order to identify different deficiencies in it. It is necessary at this point to show how Aristotle uses this analysis to

refute Socrates' claim; Charles summarises Socrates' argument and Aristotle's rebuttal thus:

> "(1) proper knowledge involves knowledge of particulars (perceptual knowledge), and (2) proper knowledge is sufficient for action, [so he] concluded that (3) proper knowledge cannot be overcome and dragged about by a slave (1145b24), and that (4) *acrasia* cannot exist (1145b25-26). Aristotle, in reply, rejects (1) by separating proper knowledge from perceptual knowledge i.e. perception of particulars (1142b25-28), since proper knowledge is concerned exclusively with universal terms. However, (3) remains correct: for it is perceptual knowledge of particulars and the last term that is the subject of attack from *acrasia* and not proper knowledge. Since Socrates arrived by chance (*sumbainen*) at the correct interim conclusion (3) by means of two mistaken but self-cancelling premises: (1) and (2), Aristotle can congratulate Socrates on this one success with mild irony; for the truth of (3) in Aristotle's view leaves open the possibility of *acrasia* (contra 4)" (Charles, 1984, p.124).

So, following a process of analysis and subdivision, Aristotle is able to identify shortcomings in a sub-type of knowledge which, if present, make *acrasia* possible—irrational emotion *can* overcome reason.

I have shown the basic mechanism in Aristotle's account by which emotion can overcome reason, but his account also explains sub-types of *acrasia*, and an outline understanding of these is also necessary before I can properly apply *acrasia* in specific choice situations.

3.3.2 Subtypes of *Acrasia*

The sub-types of *acrasia* in Aristotle's account occur in agents who are either impulsive or weak (NE 1150b19-28), and result from the effects of the feeling of pleasure or the emotion of anger (NE 1149a24-b25). An impulsive *acrates* can be distinguished from a weak *acrates* in virtue of her degree of deliberation before she acts: an impulsive agent does not deliberate before acting, whilst a weak one does so but fails "to stick to the results of the deliberation because of their affective condition" (NE 1150b20-21), i.e. because of the effect of a conflicting irrational emotion. Aristotle suggests that "quick-tempered" and "bilious" people are more likely to be impetuous: "hastiness in the one case, intensity in the other, prevent them from waiting for reason, because their disposition is to follow perceptual appearances" (NE 1150b27-29). Furthermore, it can be seen that the distinction between the impulsive and the weak *acrates* illuminates the debate described above between Aristotle and Socrates. On Aristotle's view, defended above, it should be impossible for an agent to give in to temptation "while in a state of psychologically perfect awareness that this is what one is doing" (Broadie, 2002, p.386).

However, in making the distinction between impulsive and weak types of *acrasia* Aristotle seems to acknowledge that an agent can give in to temptation in a state of perfect awareness. Broadie's explanation of this phenomenon is that even if an agent is in possession of the appropriate knowledge, it is not "making its own distinctive or typical difference to the world" (Broadie, 2002, p.386). In other words, the agent has appropriate knowledge but it is not actualised so that she can act in line with her choice rather than in line with her irrational emotions.

The second way in which *acrasia* is subdivided by Aristotle is according to whether it is the result of pleasure or anger. In the former case, an agent moves towards things that are perceived as being pleasant, and in the latter, she attacks something that is posing a threat to her. Aristotle argues that *acrasia* resulting from pleasure is more blameworthy than that resulting from anger, since reactions based on anger derive from an evaluation by the agent—syllogising of a sort—that she is under threat and that she must defend herself. Reactions based on pleasure, however, are more rudimentary: "appetite only needs reason or perception to say 'pleasant' for it to rush off to enjoy it" (NE 1149a35-36). For this reason, Aristotle says that *acrasia* resulting from pleasure is "lack of self-control without qualification" (NE 1149b19-20). However, both of the types of *acrasia* being considered here occur as a result of the same underlying process: "occurrences of temper , appetite for sex, and some things like this manifestly alter one's bodily state too, and in some people they even cause kinds of madness. Clearly, then, we should say that the state of the un-self-controlled is like these people's" (NE 1147a14-18). Now that we have Aristotle's account of *acrasia* in place, we are in a position to use it to explain potential harms to an agent who is offered an additional option.

3.3.3 *Acrasia* and Harm from an Additional Option

An agent who is offered an additional option (A), along with a pre-existing option (B), is faced with a choice, to be made after a process of deliberation, and she will normally also experience an emotional response to both options. Under the simplest construal, an *acrates* acts in line with her irrational emotions and against her choice. If, for example, her emotions take her towards the additional option (A), then this will be the option she selects. Even under this simple construal the agent can be seen to have been harmed[2] because she has gone through a process of deliberation and has chosen the option (B)

[2] I say more about what harm consists in in chapter five, where I consider how harms may be weighed against one another.

that she reasons will be best for her, but then acted against this choice and moved towards the option (A) to which she is drawn by her emotional response. So the option (A) that she selects is not the one that she reasons is best for her (B).

If, following the Aristotelian account above, we apply a more complete construal of *acrasia* to this case, the *acrates* may not have full knowledge of the particulars relating to choosing the additional option (A) or alternatively selecting the other option (B)—this may be the reason her irrational emotions overcome her choice. Alternatively, she may have knowledge of the particulars but her knowledge may not be actualized so as to make "a difference to the world" (Broadie, 2002, p.386). We should also consider the irrational emotions involved in the *acrates'* behaviour. She may have been pulled towards the option (A) because of an irrational appetitive desire, making it more pleasurable for her than (B). Alternatively, the option (B) may have been more painful for her, thereby pushing her towards the option (A). Irrational anger about (B) may also push the agent towards (A). Last, the agent may either have been impetuous or weak in this choice situation. Impetuosity, with its consequent lack of deliberation, or weakness, in which the agent follows her irrational emotions despite deliberating, may have pushed the agent towards the option (A), despite her deliberating that (B) rather than (A) is best for her. This explanation of *acrasia* in a choice situation is still a basic, unelaborated one. Consider a choice situation in which an additional option arouses strong emotions or which is novel for the agent.

In a choice situation where an additional option arouses strong emotions, these emotions, in virtue of their strength, may be more likely than weak emotions to influence how an agent acts. First in this type of case, an agent who may be volatile and prone to oscillation between *acrasia* and *enkrateia* (where she acts in line with her choice and against emotion) may be swayed towards *acrasia* if she finds herself in a situation that generates strong emotions. I have previously suggested that the agent's *hexis*, the disposition of her soul that is her character, is not completely stable, making this oscillation between character states possible. In this case, if the emotions aroused had been weaker, the agent may have acted in line with her choice. Consider a patient who has been advised to diet because she is obese and has developed maturity-onset diabetes with vascular complications e.g. sight-threatening disease of the retina. She derives pleasure from sweet deserts but, since becoming diabetic, has on most occasions been able to choose not to eat them and resist her appetite. However, she derives particular pleasure from eating sweet pastries with cream and forest-fruits. We now find

her confronted with her favourite pudding and, against her choice, she takes it from the trolley. If, however, the only available options had been different types of ice cream, for which she has less of an appetite, she may have been more likely to resist a pudding in line with her choice. So in this example being faced with an additional option that arouses strong emotions has made it more likely that the agent is harmed when she is presented with it. The harm in this case occurs because the agent's action does not reflect her reasoned decision about what is best for her.

Second, in a novel situation, an agent's ability to choose may be impaired because she does not have access to full knowledge with which to deliberate and choose about the particulars of the situation. This type of situation contrasts with alternative ones that are familiar to the agent, in which she had a good knowledge of the relevant particulars, and was able previously to deliberate and make a choice. In a novel situation, emotion may be more likely to influence the agent's action because her choice may be undermined either by lack of knowledge or because her knowledge is less likely to be actualised, making it easier for emotion to overcome choice. Consider, for example, a patient who has a chronic illness and who is offered two different treatment options for her latest relapse. Both options are equally painful but one of them offers a higher chance of a remission. The option with the higher remission rate, however, is a novel one for the agent: she has previously received the alternative treatment option. She chooses the novel treatment because of the higher remission rate, but her fear of pain overcomes this choice because her knowledge is impaired by it, or fails to be actualised, resulting in her selecting the familiar option with a lower remission rate. The agent in this example has been harmed because she receives a treatment, against her choice, that is less likely to induce remission in her illness.

Last in this account of harms to an *acratic* agent resulting from the offer of an additional option is the case in which the additional option is aligned with the agent's *acratic* tendency and may thereby encourage this tendency to be exercised. So, if we vary the case above of the patient with diabetes so that she is generally *acratic* rather than oscillating between *acrasia* and *enkrateia*, the offer of a sweet pudding of *any* type is in line with her pleasure in eating sweet foods but contrary to her reasoned choice to avoid this type of food. Since she is *acratic*, the offer of the additional option has allowed her to exercise this aspect of her character by requesting the pudding.

An alternative way in which an agent may be harmed in a choice situation where she is offered an additional option is if the additional option may be relevant in various ways to the agent's needs. So I now turn to a type of weak character relevant in this type of situation—undue self-deprecation.

3.4 Humility and Undue Self-deprecation

The second type of weak character that may result in harm to an agent who is offered an additional option is a vice, namely undue self-deprecation, which is the defect of character associated with the virtue of humility. I have previously said that the Aristotelian account of character that I am utilising incorporates the concept of different virtues, and their related character types, that are relevant in different types of situation. The choice situations that I am concerned with here are ones where the options are relevant to an agent's interests. So one plausible explanation for harm in this type of choice situation is that an agent does not have an accurate self-assessment of her worth or abilities, and thereby her interests, or reasonable concerns. If an agent is presented with an additional option and her assessment of her abilities and self-worth is accurate, then she is more likely to make a choice that is an appropriate one for her because it matches her needs. If, however, she does not have an accurate assessment of her self-worth or abilities then she may make a choice that is harmful to her because it does not match her interests. Furthermore, by their nature, certain options offered to an agent with a vice may present her with the opportunity to exercise that vice. So, in this section, I claim that the vice of undue self-deprecation is another case in which having an additional option results in a harm associated with having a weak character. This is because undue self-deprecation is one of the vices[3] related to the virtue of humility that is relevant in spheres where an agent is making an assessment of her self-worth and entitlement.

---

[3] The Aristotelian virtue of magnanimity or greatness of soul (*megalopsychia*), (NE 1123b2-12), could also be a candidate for a virtue that may help to explain harms in choice situations that are relevant to an agent's interests. This is because *megalopsychia* is a virtue in the realm of an agent's entitlement. However, after Cordner (1994) and Crisp (2006), greatness of soul, being a virtue that entails the agent seeking, having, and being entitled to *great honour*, is not suitable for my purposes in explicating potential harms to an agent who is offered an additional option. I require an account that includes entitlement to a broad range of goods and not just to great honour. Furthermore, after Sarch (2008) and Cordner (1994), the great souled person is *compelled* to actively seek great honour, and it is not relevant for the types of cases that I am considering for an agent to be compelled to selectively seek this good.

I will defend a claim, after Richards (1988), that the virtue of humility—a virtue in the area of an agent's assessment of her self-worth, abilities and entitlement—entails her having an accurate conception of her self-worth, and that there are corresponding vices of undue self-deprecation and arrogance. If an agent has an unduly self-deprecating character then she will have a conception of her self-worth as being less than that which would be justified in a fair assessment. As a result of having a lower than justified assessment of self-worth, the agent will be more prone to making harmful choices when presented with an additional option and so will experience harm.

3.4.1 Relevant Emotions for the Virtue of Humility

In order to give a complete Aristotelian account of humility it is necessary to describe the different psychological components of the virtue, which include relevant emotions. In an Aristotelian account, as I have said, a virtue must adhere to the doctrine of the mean, and features of this doctrine include feeling certain emotions associated with the virtue "when one should, at the things one should, in relation to the people one should, for the reasons one should, and in the way one should." In this general context Aristotle lists "fear, boldness, appetite, anger, pity and pleasure and distress" as relevant emotions (NE 1106b20-23). From this list, boldness, or confidence, could be a key emotion in relation to the virtue of humility. However, in the following section, I will argue that appropriate confidence constitutes a *virtue* rather than an emotion, and that this can shed light on potential harms to an agent if she is presented with an additional option. So it is necessary for me to identify an alternative emotion relevant to humility.

Aristotle gives a longer list of emotions in the Rhetoric, of which he gives proper consideration to twelve. These are: "feeling angry (*orgre*), feeling mildly (*praotes*), feeling friendly (*philia* […]), feeling hatred (*misos*), feeling afraid (*phobos*), feeling confident in the face of danger (*tharrein*), feeling disgraced (*aischune*), feeling kindly (*charin echein*), feeling pity (*eleos*), feeling righteous indignation (*nemesan*), feeling envy (*phthonos*), and feeling eagerness to match the accomplishments of others (*zelos*)" (Cooper, 1996, p.242). I claim that from this longer list, feeling disgraced (*aischune*) could be a key emotion in relation to humility and its role in explicating potential harms of additional choice. Aristotle, however, reserves a role in moral development for feeling disgraced.

Feeling disgraced could be construed as shame, and pride is the contrary of shame. So both shame and pride may be relevant to humility. People experience shame if they

assess their worth as lower than they feel it should be, and pride if they feel their worth is as high as it should be. So a virtuous agent in the sphere of humility will feel the right amounts of shame and pride in relation to her self-worth and entitlement. Since shame is reserved by Aristotle for a role in moral development, I will refer to undue lack of self-worth as the relevant emotion for undue self-deprecation. A further important aspect of undue self-deprecation is that an agent's choice (*prohairesis*) is not in a mean state.

As discussed in chapter two, an agent who is vicious does not make her choice (*prohairesis*) in a mean in relation to her. So the choice of an agent who is unduly self-deprecating will be a bad one in relation to her. For example, she will choose an option that is suitable for someone whose needs or interests are greater than her own ones. So, the agent's emotion, namely undue lack of self-worth, in combination with her bad choice, will result in her action being a bad one for her. Undue self-deprecation is a prominent feature in historical accounts of humility, so it is illuminating to consider these accounts and also necessary to reformulate them for the contemporary types of choice situation that are my main focus.

3.4.2 Reformulating Ancient Accounts of Humility

In order for me to use character types associated with the virtue of humility, in particular the vice of undue self-deprecation, to explain potential harms resulting from an additional option, humility must be taken from its historical and religious contexts and given a modern reformulation. This is necessary because, in its historical and religious formulation, humility's purpose is to serve as a reminder of man's inferiority in comparison with gods, and does not thereby allow for any virtuous state, since in its religious usage man is always inferior.

In our common understanding of humility, it is taken as an underestimation of our self-worth. This interpretation is especially prominent in religious contexts where humility is a virtue with respect to the agent's self-conception in relation to a god. Bernard of Clairvaux (1090 – 1153), cited by Richards (1988), describes religious humility in this way:

> "[I]f you examine yourself inwardly by the light of truth and without dissimulation, and judge yourself without flattery; no doubt you will be humbled in your own eyes, becoming contemptible in your own sight as a result of this true knowledge of yourself."

Whilst an agent may be justified in assessing herself as inferior in comparison with a god, and may feel shame at being "contemptible", it would not necessarily be appropriate for her to think of herself as "contemptible" in most other spheres. Richards points out that if an agent who is genuinely capable in comparison with her peers shows humility by being unduly self-deprecating, this would not be a good way for a meritorious person to think of herself: "in them, humility would be erroneous, a matter either of ignorance or self-deception" (Richards, 1988, p.253). Richards is asking for humility to be taken as an accurate self-assessment of an agent's worth and entitlement. It is important to note that the comparisons Richards is talking about are fair ones—for example, with peers. His own example is that he should not compare himself with Aristotle, in which case his work would be inferior, but compare his work with that "of contemporaries at similar institutions" (Richards, 1988, p.255).

Consider, for example, a multiple Oscar winner whose self-deprecating acceptance speech claims that the many people he mentions are entirely responsible for his success in the film and in his career. He may experience shame that he has been awarded another Oscar when his many colleagues have not been recognised. In this example we have a person of high merit whose self-assessment is that he is not meritorious. As Kupfer says: "we are dependent on fortuitous influences in our lives […] but do not need to falsely minimize the part [we play in our] success" (Kupfer, 2003, p.252). In the acting case, a virtuous agent should feel the correct amount of pride at his achievements.

There can, of course, be other permutations of self-worth and entitlement. For example, an agent may be meritorious but conceited or arrogant—she may over-estimate her abilities or worth. Consider another Oscar winner, but this time one who, when the script-writer, director and many others will have had indispensable roles in her success, does not acknowledge anyone at all in her acceptance speech. What emerges from these examples is a relationship between merit and a "humility axis", ranging from self-deprecation to conceit or arrogance. Humility is the virtue on the axis, and sits within a triadic structure that includes contrary vices of undue self-deprecation and arrogance.

Further relations between merit and humility could include non-meritorious people who are conceited or arrogant, or conversely people of the same non-meritorious ability who paint themselves less favourably than they really are. People of average abilities can also think of themselves as not being worthy of regard according with their merit. These different permutations help to bring out the key goal of humility, which is to aid an

agent in making choices and acting in ways that are consistent with a fair assessment of her worth. So, when shortlisted for an Oscar, a virtuous and meritorious actor will feel the right amounts of pride and self-worth and will not choose to reject his nomination. If he wins, his acceptance speech will give the correct amount of credit to his colleagues. In contrast, an unduly self-deprecating but meritorious actor may act in a unduly self-deprecating, i.e. vicious, way and reject the nomination. The additional option has given her the opportunity to act viciously and thereby to harm herself.

Now that we have an account of humility in place, including the relevant emotions of pride and undue lack of self-worth that make it relevant in contexts where the agent makes choices and acts in relation to her self-worth, we can see how harms may be accrued by an agent who possesses vices in relation to humility. The importance of virtue in the context of self-assessment-against-a-reasonable-standard is that "extensive false beliefs about oneself are bound to bring trouble in the long run" (Richards, 1988, p.253). If we overestimate our abilities or interests then we may not get what we think we are entitled to and may be disappointed. Conversely, if we make under-estimations then we may be harmed by not choosing or being in receipt of either what we are interested in or are entitled to. Furthermore, if an agent is in possession of either of these two vices, she will not make good judgements about appropriate goals to set for herself and to aim for.

So far in this chapter I have identified two forms of weak character that may explain harms to an agent who is presented with an additional option: first, *acrasia*, where the agent is harmed because she acts in line with her irrational emotions and not according to her choice of what is best for her, and second, undue self-deprecation, which is a vice related to the virtue of humility, which influences the agent to make choices and act according to an unfair self-assessment of her worth. A third possible form of weak character that may result in harm to an agent who is presented with an additional option is undue lack of confidence, to which I now turn.

3.5 Preliminaries for Undue Lack of Confidence

In chapter one, I analysed Dworkin's seven potential harms to an agent who is offered an additional option. One of the examples given by Dworkin to illustrate the potential harm termed "responsibility for choice" (Dworkin, G., 1988, p.67), is of a woman who has conceived a child with Down's syndrome and who is offered an abortion. In this

type of case the option presented to the woman—namely an abortion—results in her being held responsible for her actions, whether or not she has an abortion. Furthermore, the agent may be aware of a tide of opinion contrary to her own about what to do, and she may be harmed by having to take responsibility for her choice in the light of these contrary opinions[4]. Dworkin's explanation of being able to take responsibility for one's actions is that "it is a sign of moral maturity" (Dworkin, G., 1988, p.68). However, I argued in chapter one that an agent with a robust character, i.e. one who may be considered "morally mature", may still be harmed in this type of case. I also argued that a different type of harm may result to the agent in this situation, namely one that occurs if she has an undue lack of confidence in her judgements in the face of a contrary tide of opinion. An agent who has an undue lack of confidence may be harmed if she is uncomfortable, in the context of a contrary tide of opinion to her own, in taking responsibility for the choice she makes. She may feel uncomfortable about her action because of negative judgements about her choice being implied by these contrary opinions.

In order to shed light on how undue lack of confidence in judgements may result in a potential harm to an agent who is presented with an additional option, I will now give an account of confidence. I will start with a pre-theoretical account which will suggest that confidence has a triadic structure, consisting in a virtuous state of confidence and corresponding vices of undue lack of confidence and undue over-confidence. Next, I defend a claim that confidence falls within Aristotle's account of ethical virtue rather than his account of intellectual virtue since it, first, has an emotional aspect which is lacking in the account of intellectual virtue, and second, may issue in action.

3.6 A Pre-theoretical Account of Confidence

Pre-theoretically, confidence, first, could be a pro-attitude, or "evaluative response" (Blackburn, 2005, p.28) toward something. So confidence may be an attitude such as a belief that is directed toward something that is observable and assessable. Furthermore, there may be different objects of confidence with correspondingly different levels of

[4] Interestingly, it has recently been brought to public attention that the abortion rate in Iceland for women who are carrying a child with Down's syndrome is 100% (A World Without Down's Syndrome, 2016). This very high rate suggests that public opinion in Iceland is strongly against carrying and delivering children with Down's syndrome.

assessability—e.g. confidence that it will still be raining later today or confidence that a person will behave a certain way.

Second, according to its etymology, confidence may be based on trust in something. If, for example, an agent has confidence in her friend's timekeeping, then this implies that she trusts her friend to make it on-time to their get-togethers. So she thinks her friend—the object of her trust, or confidence—is trustworthy. Whether or not the agent should trust or have confidence in her friend's timekeeping may be determined by deliberation, perhaps about her timekeeping for previous get-togethers (Hardin, 2005, p.926). Following this process of deliberation, the agent may perhaps lack confidence in her friend or alternatively may be over-confident in her. For example, she may believe that her friend will never turn up in time for their meetings, or she may believe that she will always be on-time, without fail.

In the types of case that are my particular focus, the agent in a choice situation may have an undue lack of confidence in her values or beliefs—she may make a negative evaluation of them. She may, for example, have an undue lack of confidence because there is some reason for her to doubt her beliefs or values. Since an agent's values are an aspect of her mind, her confidence in her values will form after a process of introspection. Introspection is the process of learning about, or paying attention to, one's mind, including its states and processes. Introspection has certain conditions imposed on it, such as effort (it is not constant, effortless and automatic), temporal proximity, being first person, being detected in a direct or immediate way, and being a form of mentality (Schwitzgebel, 2016).

So, confidence in one's values may be a declarative attitude resulting after a process of perception and introspection about relevant mental states, and from deliberation about those mental states. Furthermore, confidence has a triadic structure, since an agent may either lack confidence in her values, be appropriately confident in her values, or be over-confident in her values.

3.7 Confidence as an Ethical Virtue and not an Intellectual Virtue

According to my pre-theoretical account of confidence, it has a triadic structure—an agent can be under-confident, over-confident, or appropriately confident. We have already seen in chapter two that the ethical virtues have a triadic structure with two

corresponding vices of excess and deficiency, but the intellectual virtues may also have a triadic structure (Gottlieb, 2009). So confidence could either be regarded as an ethical or an intellectual virtue. This distinction is relevant to my thesis because intellectual and ethical virtues have differences in their component parts and in the ways that they are acquired and function (NE II.1 and NE VI). Importantly, however, confidence may have an emotional component that is missing from intellectual virtue (NE 1139b12-14), and this takes it away from intellectual virtue and into the domain of ethical virtue.

In order to illustrate confidence being associated with an emotional component, and thereby being an ethical virtue, let us consider a cellist who is due to perform in a public recital. Unfortunately she lacks confidence in her playing ability since she sometimes gets "bow-shake", and this results in her being frightened that she will get bow-shake during the recital. An agent may feel pleasure or pain depending on how confident she feels, and if she is virtuously confident then she will feel the correct amounts of pleasure or pain. If, however, as in the case of the cellist, she is very lacking in confidence, this may be associated with a negative, painful emotion such as fear.

In this case, the cellist is aware of how she has played in previous performances, in front of different audiences and at different venues. In the light of truths about previous performances etc. there are different possible ways in which she may perform in her next recital, and she may deliberate about these different possibilities. Aristotle tells us that we do not deliberate about things that cannot be otherwise but "[w]hat we do deliberate about are the things that depend on us and are doable; and these are in fact what is left once we have been through the rest" (NE 1112a31-2). Furthermore, "[d]eliberation […] occurs where things happen in a certain way for the most part, but where it is unclear how they will in fact fall out; and where the outcome is indeterminate" (NE 1112b7-9). So suitable objects for deliberation by the cellist would include the different qualities of her previous performances and, in the light of these, the potential qualities of future performances since they "depend on us" and are "indeterminate". So, by deliberating about past and future performances, the cellist attempts to situate her confidence appropriately on a continuum from severe lack of confidence at one end to extreme over-confidence at the other.

Somewhere along the confidence continuum there will be an appropriate degree of confidence, reflecting a fair assessment of all the relevant aspects of the cellist's playing. Unfortunately, in the case above, the outcome of the cellist's deliberation is that she is

much less confident about her upcoming performance than she should be based on her abilities. She thereby has the vice of undue lack of confidence in her playing.

The cellist who lacks in confidence may seek the help of a teacher or performance psychologist. Under supervision and after re-deliberating about her abilities, she may consider all the practice she has done and the improvements she has recently made in her bow-technique and reach a revised, proper state of confidence in her playing. Furthermore, as a result of this true state of confidence in her playing, she may play confidently and without bow-shake—an action that has issued from this confidence. This last feature of confidence, namely that it issues in action, also places it within an account of ethical virtue; intellectual virtue does not issue in action. If an agent's deliberations about an object of confidence go well and she is virtuously confident, she is described by Aristotle as being practically wise: "we […] call those in a specific field wise if they succeed in calculating well towards some specific worthy end on matters where no exact technique applies" (NE 1140a29-30).

So far in this section I have defended an account of confidence as an ethical virtue. Confidence about things, then, conforms to a triadic structure, is the outcome of a deliberative process with associated emotions, and issues in action. My last claim in this section, to which I now turn, is that a sub-type of confidence—undue lack of confidence in one's judgements—may result in harm to an agent in a choice situation.

3.7.1.1 Confidence in One's Judgements

An agent may be harmed when she is presented with an additional option if she finds it difficult to take responsibility for her choice because she has an undue lack of confidence in her judgements based on her values[5]. Returning to Dworkin's example of "responsibility for choice" (Dworkin, G., 1988, pp.67-68), it is possible to see this harm being accrued by a woman who is able to identify through prenatal testing that her unborn child will have Down's syndrome, and who has the additional option of an abortion. If the woman wishes to refuse an abortion and keep her child, she may be harmed by having to take responsibility for a decision not to have an abortion in the face of other people who do not feel that her choice is the correct one. I argued in

[5] An alternative way in which an agent may be harmed in a choice situation if she has an undue lack of confidence in her judgements is if she finds it difficult to hold on to her values in the face of a contrary tide of opinion and consequently acts against her values.

chapter one that an agent with a robust character may be harmed in this situation as a result of the graveness of the decision being made. That is to say, even if the woman had been confident in her judgements, she would still have been harmed by having to take responsibility for her decision.

I claim that harm can also result to the woman in this case if she has the vice of undue lack of confidence in her judgements. An agent's judgements are "indeterminate" (NE 1112b7-9), and may thereby be the subject of her deliberation. She will also experience emotions in association with her confidence in her judgements. If, then, her deliberation about her judgements does not result in her being properly confident in them, she may experience harm through not being able to take responsibility for her choice.

In this section I have explained the third of three types of weak character—the vice of undue lack of confidence in one's judgements—that may result in harm to an agent. When faced with an additional option, the agent is presented with the opportunity to exercise this vice and may thereby be harmed as a result of finding it difficult to take responsibility for her choice.

3.8 Conclusion

In this chapter, I have defended accounts of three different types of weak character derived from the Aristotelian account of character made plausible in chapter two. I argue that an agent who has any of these three types of weak character may be harmed in a situation where she is presented with an additional option. First, an *acratic* agent may be harmed if she is offered an additional option that presents her with an opportunity to exercise her *acrasia*. Furthermore, she is more likely to be harmed in this type of situation if it either arouses strong emotions or is a novel one to her; in both of these two sub-cases the agent's choice is less likely to overcome her irrational emotion. The harm to the *acratic* agent occurs because she requests an option that she has reasoned is not the best one for her. So if the additional option had not been presented to her she would not have been harmed.

Second, an agent who is unduly self-deprecating may be harmed in a choice situation if she requests an option that is not suitable to meet her needs as identified in a fair assessment. Last, an agent who has an undue lack of confidence in her judgements may be harmed as a result of finding it difficult to take responsibility for her choice. The harm is more likely in this case if the agent is aware of other people in the choice

situation whose values differ from her own. So the agent here is influenced by awareness of an aspect of the context of choice which I now defend in full in chapter four. In particular, I will defend an account of the normative features of context of choice and how these explain potential harm to an agent who is presented with an additional option.

Chapter 4. The Context of Choice

4.1 Abstract

In this chapter, I will develop my argument from chapter one that there is a second aspect of a choice situation (after certain types of weak character) that is relevant to the harm that may be incurred by agents who are offered an additional option. This aspect is the context in which an agent chooses, which I term the context of choice. Drawing on the etymology of the word *context*, I shall limit the scope of the context of choice to features of the choice situation that are *woven into* an agent's decision-making because of their salience to her. I will also argue for an account of the context of choice using resources from the literature on, first, bounded rationality (BR) and second, adaptive preferences (AP), since both these concepts have as their focus different types of choice situation and the ways in which these can affect an agent. The theories of BR and AP show how a choice situation may be altered in harmful or beneficial ways as a result of alterations in the salience to the agent of aspects of this situation. Last, I will use a medical analogy to illustrate normative features of the context, namely ones that either should or should not be salient to an agent.

My account of the context choice is that it consists in features of a choice situation, first, that are salient to an agent and either should or should not be salient to her, second, that should be salient but are not salient to her, and third, that interact with other features in the choice situation to alter their salience to her. The normative features of the context of choice help to explain how it can influence an agent so that she may be harmed if she is presented with an additional option. Features of the choice situation that should be salient to the agent are those that enable her to make a rational decision. So an agent may be harmed in a choice situation if features of it either that should be salient are not salient to her, or that should not be salient are salient to her. My position in this chapter goes beyond those of Dworkin, G. (1988) and Velleman (2015), because it explains how contextual influences on an agent in a choice situation, i.e. the context of choice, may result in a further type of harm to her when she is offered an additional option.

4.2 Introduction

My pre-theoretical account in chapter one of the context of choice consisted in features of the choice situation external to an agent of which she is aware and which may

thereby affect her decision-making. I sub-divided these features into two: first, into various properties of other people in the choice situation, such as their values, intentions and behaviours, and second into physical features of the choice situation such as their value to the agent, their complexity, and the broader environment in which they are embedded. I used this pre-theoretical account to defend a claim that none of the nine harms to an agent identified by Dworkin and Velleman are dependent on adverse features of the context of choice. Furthermore, adverse features of the context of choice can explain additional harms to the agent that are distinct from Dworkin's and Velleman's harms. My pre-theoretical account of the context of choice was a descriptive one that lacked both a definition and an explanation of its normative aspects. So this chapter will both define the context of choice and explain how it may influence an agent and potentially result in harm to her as a result of its normative features.

In order to define the context of choice and bring out its normative aspects, first, I motivate my line of argument by briefly examining evidence from the social sciences on the effects on an agent's choice of the environment in which she is choosing. Next I revisit selected examples from Dworkin and Velleman that identify features of a choice situation that should be included in an account of the context of choice. As the scope of the context of choice could be unmanageably large, I show how it can be limited to features of a choice situation that are woven into the agent's decision-making. I then develop an account of the importance of changes in salience of features of the context of choice to an agent in an altered choice situation, first, by drawing on two relevant areas of literature—namely bounded rationality and adaptive preferences—and second, through a medical analogy.

The literatures on BR and AP are relevant for my purposes since they both analyse situations in which an agent undergoes an important psychological change as a result of aspects of her context. First, I divide BR into negative and positive claims and argue that the negative claim shows how an agent's decision-making may be adversely influenced in a choice situation as a result of changes in the salience to her of features of this situation. The positive claim in BR shows how it is possible for an agent to improve her decision-making through the use of heuristics, and shows how this may be achieved by rendering salient to her features of the choice situation that should be salient. The changes in salience to an agent of features of a choice situation described in the BR literature suggest that there are normative aspects of the context of choice which are relevant to potential harms to an agent in a choice situation. Second, I argue that the

literature on AP also helps to illuminate the context of choice. This is because it describes choice situations where the range or quality of options on offer is altered and, as a result of the agent adapting her preferences, the salience to her of features of the choice situation are also altered.

I then draw on a medical analogy to defend a claim that the context of choice consists in aspects of a choice situation, first, that are salient to the agent and either should or should not be salient to her, second, that are not salient to the agent and either should or should not be salient to her, and third, which interact with other aspects of the choice situation to alter their salience to the agent. The aspects of the choice situation that should or should not be salient to the agent are those that enable her to make rational choices. In the medical analogy, aspects of the context of diagnosis, which is comparable to the context of choice, may either help or hinder the doctor in making the correct diagnosis. The medical analogy shows, then, how there is potential for an agent to be harmed in a choice situation if she finds salient some feature of the choice situation that she should not find salient or does not find salient a feature of the choice situation that she should find salient. Last, the medical analogy also shows how it may be possible to generalise about what may make a good or bad context of choice.

Having established that features of the context of choice that should be salient to an agent are those that enable her to make a rational decision, I consider whether it may be possible to configure the context of choice in such a way that she is not harmed in making her decision. A "good" context of choice, then, would be one in which the salience or otherwise of its features are not altered in a way that prevents the agent from making a rational choice.

The chapter concludes with the claim that alterations to the context of choice, for example through the introduction of an additional option, may harm an agent in a choice situation, due to changes in what is salient or otherwise to her when she is making her choice.

4.3 Motivation for Developing an Account of the Context of Choice

I have argued that an agent may potentially be harmed as a result of features of her character if she is offered an additional option. However, this may not be the only way in which an agent can be harmed in a choice situation. A large body of evidence from the social sciences describes effects on decision-making of variations in the way options

are offered—for example, by "framing" (Tversky and Kahneman, 1986), "decoys" (Tversky and Simonson, 1993) or "positioning" (Thaler and Sunstein, 2008). These effects appear to originate in alterations in the context in which the agent is choosing, which is a feature of the choice situation that is also outside the agent's control. Furthermore, it is possible for these contextual aspects of the choice situation to be open to manipulation by other agents. Since features of the context may influence the choice the agent makes, they may potentially also result in harm to her. So, a deeper understanding of what I term the context of choice is important in understanding further ways in which an agent may be harmed when presented with an additional option.

4.4 Preliminary Examples

As a first step in an account of the context of choice it is helpful to return to examples of choice situations used by Dworkin and Velleman in their arguments about potential harms from additional choice. Two such choice situations are shopping for a shirt (Scitovsky 1976, quoted by Dworkin, 1988, p.54) and receiving an invitation to a dinner party (Velleman, 2015). I argued in chapter one that harms accruing to an agent who has a type of weak character are distinct from the harms described by Dworkin and Velleman. In the first situation, an agent is shopping for a shirt and is hypothetically moved from a simple situation where there are few options to one where she is faced by numerous options. I have argued that the agent may potentially be harmed in this situation as a result of her having certain types of weak character such as undue lack of confidence in judgements which, for example, renders her more susceptible to peer pressure. In the second example, an agent is given the additional option of attending a dinner party. She may be harmed by receiving the invitation because she no longer has the default option of not attending, and must now choose either to attend or not to attend, but she may be additionally harmed as a result of relevant features of her character—such as undue lack of confidence in judgements or undue self-deprecation.

I will now analyse these two situations in respect of the contextual aspects of the choice situation that also explain how the agent may be harmed. In the first of the two, contextual aspects that may explain potential harms to an agent who is offered an additional option are the number of shirts (e.g. too many), the range of different colours and styles (e.g. too wide), the shirts' values (perhaps a wide range), and the way in which they are arranged (e.g. confusingly). In the second case, contextual features include the host's valuation of the invitation (e.g. high), whether or not the agent likes

the host (e.g. she does not), the status of the host (e.g. someone of importance), the travel required (e.g. a great distance) or the menu for the meal (e.g. mainly game). These non-exhaustive lists of contextual features for each case may be part of what I term the context of choice. So, pre-theoretically, the context of choice could include *inter alia* various features of a choice situation including the number of options on offer, the way the options are arranged, the valuation of the options to the agent and others, and the complexity of the options. A suitable account of the context of choice that can help explain the potential harms to agents who are offered an additional option is now required.

A first problem in defining the context of choice is how to limit it. After all, the agent is making her choice in a broader context than one consisting merely in multiple options and other peoples' opinions. For example, in the second case, the agent may be aware that Jupiter is especially close to the Earth on the evening of the meal, but should this and other astronomical facts be part of the context of choice? The broader context in which an agent is choosing could be unmanageably large if it were to include everything in the known universe, so it is clearly necessary to limit the context of choice.

4.5 Limiting the Context of Choice

An agent making a choice does not do so in isolation; she chooses within a context that includes a range of situational features. So the context of choice could potentially encompass spatial and temporal aspects of everything sensate in the universe, akin to the "blooming, buzzing confusion" that James suggests a baby experiences without a properly formed perceptual apparatus (James, 1981, p.488). Furthermore, the context of choice could also include everything that the agent imagines, including things in the past and future, so perceived risks and benefits could be part of the context of choice. Clearly a context of choice this broad would be intractable for an agent.

A potential starting point for a workable account of the context of choice, in which it is not unmanageably large, is from the etymology of *context*. From the Latin for "woven into", context suggests that what is needed for an account of the context of choice is an understanding of what is woven into choice. Textiles are made by weaving a transverse weft through a longditudinal warp. So if the choice made by the agent is the warp in a textile, then the context of choice may be the weft. The weft in a textile "influences" the warp by being woven into it. Furthermore, if the weft is pulled on or otherwise

manipulated, then this may have an effect on the warp. Developing this analogy, the context of choice will interact in an important way with the agent's choosing. Altering (or tugging on) the context of choice (weft) will have an effect on, or influence, the choice (warp). So this analogy suggests that what we may be looking for in the context of choice are features of the choice situation that influence the agent's choice by being woven into it.

A benefit of the weaving analogy is that it can show us how the scope of the context of choice may be restricted. When an agent chooses, she first perceives certain features of the choice situation as salient—in other words they leap out at her—and then she actively uses them to influence her choice. Consequently, these salient features, and not those that remain in the background, are in effect woven into her decision making and choice. The corollary of this is that the agent is not assailed to various degrees by all the other aspects of the choice situation. So, for example, Jupiter, Venus or Mars, may not be part of the context of choice for the dinner party invitee.

If, however, the invitee *is* interested in the positions of planets and stars, and these aspects of the context of choice are salient to her, then they may by her own lights be relevant to her making a rational choice. To illustrate the relevance of salience to an agent making a choice, consider a woman, Jane, walking home on a clear night after an evening at the cinema. She remembers, as she is walking, that tomorrow is the day the bins are emptied, and wonders if she should either put them out before she goes to bed or first thing in the morning. As she is walking she is aware that she can see Jupiter. Perhaps Jupiter being visible ought not to be salient in order for her to make a rational decision about whether or not to put the bins out, since for most people it is probably not relevant to this decision. However, Jane is interested in astrology, so being able to see the position of Jupiter may influence her in her choice about when to put out the bins.

In a variation to this astrology case, it is cloudy as Jane comes back from the cinema, and, despite her interest in astrology, Jupiter is not salient to her as it is not visible—it does not leap out at her. By Jane's own lights, however, the position of Jupiter should be salient to her—it should weave itself into her decision-making in virtue of being part of the context of choice. In this variation, the context of choice may include aspects of the choice situation that should be salient to Jane but are not. On the other hand, the International Space Station, which is brighter than Jupiter on a clear night, is not

important to Jane by her own lights and thereby unimportant to her decision about when to put the bins out, so it does not matter to her that she cannot see it. The cloudy variation, then, suggests that there may be additional features of the context of choice in virtue of their relevance to the agent by her own lights. First, there may be features of the context of choice that should be salient to the agent by her own lights but which are not salient to her, and second, there may be features of the context of choice that are not salient to her and should not be salient to her by her own lights. These two additional aspects of the context of choice are agent relative ones, but there are also non-agent relative aspects of the context of choice.

In a further variation of the dust-bins case there are now thunder-clouds overhead, making it very likely that it will rain heavily. Jane's bins have ill-fitting lids and leak in the rain. Furthermore, she has to put them out on a surface that is prone to flooding in heavy rain. So the thunder clouds in this case should be salient to Jane—they are a feature of the context of choice which, if she finds them salient, will enable her to make a rational decision about when to put the bins out. The thunder clouds should also be salient to any other agent in the same type of situation, because rubbish being washed down the street is not healthy. This last variation brings out a further potential aspect of the context of choice: namely features of the choice situation that should be salient to any agent in a similar situation.

In this section I have used a linguistic cue to limit the context of choice to features of a choice situation that are woven into an agent's decision-making. I have identified that these features are those that have, or should have, salience to an agent in that they enable her to make a rational choice. These features—that for various reasons either should or should not be salient to the agent—suggest that there may be a normative conception of the context of choice which I fully develop in a later section.

At the beginning of this section I mentioned empirical research from the Social Sciences that supports the claim that the context in which the agent chooses is relevant to the choices she makes. The claimed effects in these cases are due to "framing" (Tversky and Kahneman, 1986), "decoys" (Tversky and Simonson, 1993) and "positioning" (Thaler and Sunstein, 2008). I now analyse the choice situations in these cases and draw out key features which further develop my account of the context of choice. I then turn to the broader concept underlying these cases—that of bounded rationality.

4.6 Cases from the Social Sciences Showing Effects of Context

The first of the three empirical cases from the Social Sciences literature that I examine shows what is termed the "framing effect". Consider this hypothetical example of medical decision-making from McNeil *et al*. quoted in Tversky and Kahneman (1986). Respondents in a study were asked to state their preferred treatment option.

> "Problem 1 (Survival frame)
>
> Surgery: Of 100 people having surgery 90 live through the post-operative period, 68 are alive at the end of the first year and 34 are alive at the end of five years.
>
> Radiation Therapy: Of 100 people having radiation therapy all live through the treatment, 77 are alive at the end of one year and 22 are alive at the end of five years.
>
> Problem 1 (Mortality frame)
>
> Surgery: Of 100 people having surgery 10 die during surgery or the post-operative period, 32 die by the end of the first year and 66 die by the end of five years.
>
> Radiation Therapy: Of 100 people having radiation therapy, none die during treatment, 23 die by the end of one year and 78 die by the end of five years" (Tversky and Kahneman, 1986, S254).

In this choice situation the information about each treatment presented to the research subjects was factually identical but "framed" differently. This was achieved by formulating the information in one case in terms of the numbers of people who *survive* the different interventions and in the other case in terms of the numbers of people who die (*mortality*) following the interventions. Tversky and Kahneman (1986, S255) describe the results of this framing thus:

> "The inconsequential difference in formulation produced a marked effect. The overall percentage of respondents who favoured radiation therapy rose from 18% in the survival frame (N = 247) to 44% in the mortality frame (N = 336). The advantage of radiation therapy over surgery evidently looms larger when stated as a reduction of the risk of immediate death from 10% to 0% rather than as an increase from 90% to 100% in the rate of survival. The framing effect was not smaller for experienced physicians or for statistically sophisticated business students."

Surgery comes out as the preferred option in both frames, probably because of the superior number of patients alive at five years, but radiation therapy, when presented within the mortality frame, increases in popularity. This is presumably because of the favourable comparison between radiation and surgery in the post-operative period, if the comparison is assessed unreflectively by the respondent. So in terms of my account so

far of the context of choice, which has features of a choice situation salient to the agent within that situation, we can identify the following key information. Since there are no arithmetical differences between the information in the two frames, the agent should not find information about radiation therapy in the mortality frame to be more salient than the same information when it is presented in the survival frame. Consequently, she should not be more likely to choose radiation therapy in the mortality frame than she is in the survival frame, since the outcomes in both frames are identical.[1] So a change in context—in this case, the way in which the information is framed—has had an effect on the agent's decision-making in virtue of altering the salience to her of aspects of the choice situation.

The Decoy Effect (Tversky and Simonson, 1993) is the second type of case in the empirical literature that shows the effect on an agent's choice of the context in which she is choosing. The Decoy Effect occurs in two types of choice situation where agents are presented with two options, A and B, and where A should be preferable to most people. In the first instance, A and B are the only options on offer, and most agents choose A. In the second instance, A and B are offered with a third option, C, where C compares less favourably with B, but cannot easily be compared with A. When the agents come to choose from A, B and C, fewer of them prefer A in comparison with B. The Decoy Effect has been confirmed empirically by Tversky and Simonson:

> "One group (n = 106) was offered a choice between $6 and an elegant Cross pen. The pen was selected by 36% of the subjects and the remaining 64% chose the cash. A second group (n = 115) was given a choice among three options; $6 in cash, the same Cross pen, and a second less attractive pen. The second pen, we suggest, is dominated by the first pen but not by the cash. Indeed, only 2% of the subjects chose the less attractive pen, but its presence increased the percentage of subjects who chose the Cross pen from 36% to 46%, contrary to regularity" (Tversky and Simonson, 1993, p.1182).

Tversky and Simonson offer the following explanation for the Decoy Effect: "context effects, in perception as well as in choice, provides numerous examples in which people err by complicating rather than by simplifying the task; they often perform unnecessary computations and attend to irrelevant aspects of the situation under study. [...] [T]he

---

[1] Kahneman explains the effects of framing by attributing them to so called "System 1" mental processes, which he defines by their tendency "automatically and quickly, with little or no effort, and no sense of voluntary control" to "generate impressions, feelings, and inclinations" (Kahneman 2011, p.105.) As Kahneman also says: "System 1 […] is rarely indifferent to emotional words: mortality is bad, survival is good, and 90% survival sounds encouraging whereas 10% mortality is frightening" (Kahneman 2011, p.367).

easiest way to decide which of two options is preferable is to compare them directly and ignore the other options" (Tversky and Simonson, 1993, p.1188).

There are two key points that the Decoy Effect has established. First, if an agent is presented with an additional option then this may affect the salience to her of the other options. In other words, an alteration to the context of choice by adding a third option modified the salience to the agent of other features of the context of choice. Second, the additional option had a potentially harmful effect on the agent's choosing as a result of the alteration in salience to her of aspects of the pre-existing options. So, the context of choice in the decoy cases includes features that should be salient to the agent—the difference in value between A (the cash) and B (the Cross pen)—and features that ought not to be salient but which are—the favourable comparison between B (the Cross pen) and C (the less attractive pen).

The third type of case in which contextual features have an effect on the choice the agent makes is one where the objects in the choice situation are positioned differently. Thaler and Sunstein give a paradigm example of this type of case in "Nudge" (2008). In their example, the spatial positioning of the features of the choice situation are manipulated and the consequent effects on decision-making are exploited to influence people to make so-called beneficial choices. A hypothetical director of school food services, Carolyn, discovers that the choices of food made by her students can be influenced by the positioning of the food.

> "In some schools the deserts were placed first, in others last, in still others in a separate line. The location of various food items was varied from one school to another. In some schools the French fries, but in others the carrot sticks, were at eye level. […] Simply by rearranging the cafeteria, Carolyn was able to increase or decrease the consumption of food items by as much as 25%. Carolyn learned a big lesson: school children, like adults, can be greatly influenced by small changes in the context" (Thaler and Sunstein, 2008, p.1).

A key feature of choice situations such as this one is that the range of options on offer is not altered: none of the options available to the students is "closed off" or rendered "appreciably more costly" (Hausman and Welch, 2010, p.136). Despite this, in one variant of this example of "nudging", an agent who by her own lights has a preference for chips over carrot sticks may not buy the chips if they are moved in the cafeteria so that they are less prominent than the carrot sticks. So altering the positioning of the options may have an effect on the salience to the agent of those options: in the variant above, the chips do not leap out at the agent, but the carrot sticks do.

An analysis of these three types of case in the Social Sciences literature draws out the following aspects of the context of choice. First, from framing, the context of choice includes features of a choice situation that can interact in a significant way with other features of a choice situation to alter the latter's salience to the agent. Interaction here means altering the salience of existing options by virtue of the way they are presented. Second, from the Decoy Effect, features of a choice situation may become salient to an agent who is presented with an additional option when they should not be salient to her. Last, from Nudge, alterations in the presentation of features in a choice situation may also alter the salience to an agent of those features. Nudge also draws attention to normative aspects of the context of choice—perhaps an agent who is morbidly obese and who by her own lights prefers chips to carrot sticks should find the carrot sticks more salient in order to improve her health[2].

Each of these three cases from the Social Sciences claims to give evidence supporting the theory of BR, so I now turn to BR itself. What I seek to establish about the context of choice through a discussion of BR, is that alterations may be made to the context which in turn alter the salience to an agent of different aspects of a choice situation. These alterations may have either harmful or beneficial effects on the agent.

4.7 Bounded Rationality

The theory of bounded rationality (BR) is a response to an account in economics of so-called ideal rationality. The economic account assumes that an agent in specific choice situations should use certain intellectual resources, such as Expected Utility Theory and Bayes' Theorem (on which I will expand later), in order to deliberate and make decisions in an optimally rational way. The first account of the concept of BR was put forward by Simon (1957), in response to this supposedly rational man, or *homo economicus,* who he described as having

> "knowledge of the relevant aspects of his environment which, if not absolutely complete, is at least impressively clear and voluminous. He is also to have a well-organised and stable system of preferences, and a skill in computation that enables him to calculate, for the alternative courses of action that are available to him, which of these will permit him to reach the highest attainable point on his preference scale" (Simon, 1957, p.241).

---

[2] The UK government jointly own the Behavioural Insights Team (2016) which has as one of its objectives to enable "people to make better choices for themselves."

Unsurprisingly, the abilities of *homo economicus* as described by Simon (above) are unattainable by most agents. First, there is no explanation of how *homo economicus* may come to have an "impressively clear and voluminous" knowledge of the relevant aspects of his environment, especially if the environment is new to him. Second, he, as many others, may not have a "well-organised and stable system of preferences", not least if he is also in a new environment and facing new options. Last, the requirements for *homo economicus's* calculating abilities imply that the situations he faces are ones in which there is in fact a way of calculating the best action. Many choice situations, however, such as choosing a life-partner, are ones where there is no suitable metric to facilitate a calculation. Furthermore, accounts of so-called ideal rationality, e.g. Expected Utility Theory, appear to fit the economist's narrow conception of a rational man as being one who seeks to maximise his own self-interest.

The account of BR that emerges accommodates the ways in which ordinary people normally fail to apply the principles of ideal rationality in order to make optimal choices. BR consists in a negative account of the ways in which an agent may be constrained if she tries to follow the principle of ideal rationality, and a positive account of second best strategies, or heuristics, that an agent who has limited rational resources can adopt to arrive at the most rational decision that she is capable of. Importantly, both the negative and positive claims in BR explain how alterations in the salience to an agent of aspects of the choice situation may respectively have harmful or beneficial effects on her.

In the following sections I will relate the empirical research that clams to support the negative claim in BR, showing impaired rational decision-making in specific situations, such as in the framing and decoy cases above. I will also outline armchair arguments about how rationality is limited, before turning in more detail to the negative and positive claims in BR.

4.7.1 Empirical Evidence for Bounded Rationality

The negative claim in BR has repeatedly been demonstrated in empirical studies of decision-making. Two such studies are, first, an investigation of Allais' paradox, which breaches Expected Utility Theory (Tversky and Kahneman, 1992), and, second, a further study by Kahneman and Tversky on "The Psychology of Prediction", which shows that agents often disregard Bayes' Theorem (1973).

First, Allais' paradox examines a situation in which agents are presented with two pairs of complex choices, first, between A and B and second, between C and D. Furthermore, the differences between A and B, and C and D are evaluatively equivalent, which can only be recognised when carefully analysed as if by *homo economicus*. Despite these evaluative equivalences, 82% of subjects in the study chose B and 83% chose C. In this case, the normal subjects were unable to properly evaluate the different options using basic arithmetic and so failed to live up to the standards of *homo economicus*—exhibiting instead bounded rationality.

Second, Kahneman and Tversky (1973) analysed whether agents appropriately use Bayes' Theorem, which states that "the probability of a hypothesis depends on the prior probability of the hypothesis" (Grüne-Yanoff, 2007, p.539). Subjects in one study were given a numerical account of members of a sample population which consists of two characteristic groups—e.g. nuclear scientists and social workers. They were then given descriptions of individuals taken from the sample population and asked which of the two characteristic groups the individuals come from. The descriptions were either carefully worded in order to bias the subject towards identifying them as belonging to one or other of the groups, or worded to be neutral. In a second part of the study the overall proportions of the characteristic groups were reversed. Despite this reversal of the proportions, subjects normally continued to assess the likelihood that an individual description matched one of the two groups according to the stereotypical or neutral description itself, and not according to the proportions. In other words, they failed to use prior probabilities when assessing the identity of a group member from a description.

The studies on Allais' paradox and Bayes' Theorem (and many others) claim to give evidential support for BR. In doing so they depend on assumptions that the rational way of approaching the relevant exercise is, for example, by using Bayes' Theorem or Expected Utility Maximisation. It is beyond the scope of this dissertation to disprove either Bayes Theorem or Expected Utility Maximisation and the role they should have in decision-making. Instead, I continue by describing two armchair arguments in support of BR's negative claim, before showing how this negative project is relevant to the context of choice.

4.7.2 Armchair Arguments for Bounded Rationality

One argument for the negative claim in BR is based on the idea that "many models of rational inference [i.e. according with the concept of *homo economicus*] treat the mind as a Laplacean Demon, equipped with unlimited time, knowledge and computational might" (Gigerenzer and Goldstein, 1996, p.650). Laplace's demon is able to calculate the probability of all future events because it has perfect knowledge of all past states of the universe and also knows how these states cause future states. One weakness inherent in this comparison is that the scope of resources necessary for rational inference in an everyday choice situation is not as broad as it is for the Demon when it computes all future states of the universe. The capabilities of Laplace's demon greatly exceed those of *homo economicus* who, despite having effective "skill[s] in computation" has incomplete but "clear and voluminous" knowledge (Simon, 1957, p.241). However, as I have said, in most choice situations it is still beyond human capabilities both to have sufficient knowledge and to calculate all of the various relevant consequences. This, after all, is a prominent objection to Consequentialism.

A further argument in support of the negative claim in BR is from Zermelo (1913)[3], who claimed that chess would be a determinate game if either of the players had sufficient computational abilities. None of us consider that chess is determinate—if it were then there would not be any point in playing it as a game—so rationality is "bounded" in the sphere of chess. The exact detail of Zermelo's claim is disputed (Schwalbe and Walker, 2001), but, even accepting this dispute, people's rational ability to play chess falls within a range. People lower down on the range have to choose between, on the one hand, increasing computational costs if they try to work out all the permutations several moves or more ahead, and on the other hand, a greater chance of losing the game if they expend less of their resources on making such computations. Players who conform to the standard set by *homo economicus* are not faced with this choice because they have copious knowledge and sufficient computational skills to reach the "highest attainable point on [their] preference scale" (Simon, 1957, p.241). Even they, however, would be beaten in a game of chess by Laplace's demon. So the point that is established here is that humans can only approach a state of ideal rationality: human rationality is indeed "bounded". An important part of the negative, "bounded" component of BR for my purposes, is the explanations given for the

[3] I am grateful to Christina Nick for translating this paper.

apparent failings of agents in particular choice situations. One such explanation is given by Kahneman (2012).

The explanation given by Kahneman (2012) for the cognitive errors that agents may make in certain choice situations is based on his concept of "fast, System 1" and "slow, System 2" thought processes. Agents, so the account goes, have a tendency to make decisions using error-prone "fast" cognitive processes rather than more accurate "slow" ones. Fast processes are intuitive, automatic, credulous, they "frame decision problems narrowly", and "respond to losses more than to gains" (Kahneman, 2012, p.105). Slow processes, however, are more likely to be accurate since they are "conscious, slow, controlled, deliberate, effortful, statistical, [and] suspicious" (Shleifer, 2012, p.3). So in the experiment on Allais' paradox, agents frequently rely on a rapid, intuitive assessment of the pairs of options on offer and falsely assess the differences between them as being non-equivalent. In the "prediction" experiment, also mentioned above, agents rush to label a group member based on a quick, intuitive assessment of the description, but without taking any account, utilising System 2 processes, of the probabilities inferred by the overall make-up of the group. If, however, the subjects in both experiments had engaged System 2 processes, they would have identified either the non-equivalence of the options, or taken account of the importance of prior probability in assessing the identity of the group members from their descriptions.

However, there are objections to Kahneman's account. In his review of Kahneman's "Thinking Fast and Slow", Schleifer claims that BR "is very different from […] System 1" since BR predicts that people fail to solve hard problems, whereas System 1 explains why people "get utterly trivial problems wrong because they don't think about them in the right way" (Shleifer, 2012, p.4). In response to Schleifer, even trivial problems sometimes need rational solutions; an agent may not be able to form intuitions about a novel situation where she does not recognise any of the relevant features. So an account, even in more simple situations, that explains a failure of agents to fully use rational processes, as Kahneman's does, is an account of how rationality is "bounded". Furthermore, as Schleifer admits, decision errors may still possibly be attributable to System 2 failures—i.e. failures of rational processes—though errors may also be caused either by the agent using System 1 processes or a combination of System 2 errors and

System 1 processes. All these potential causes of errors in decision-making help to explain why agents are not fully rational, and so should be part of BR[4].

In the sections above, I have analysed the negative claim of the theory of BR—that agents are non-ideally rational and thereby make less than optimal choices. I have also analysed empirical studies in support of this claim. These analyses help me to defend my main claim, that the theory of BR sheds light on the context of choice. The errors in decision-making made by an agent in the type of cases that I have outlined occur as a result of her finding salient features of the choice situation that she should not find salient due to alterations in some features of that situation. Kahneman, however, explains the errors an agent makes in these situations utilising his concept of System 1 and System 2 thought processes. His system 1 processes are fast, "frame decision problems narrowly" and also "respond to losses more than to gains" (Kahneman, 2012, p.105). I have already discussed framing in a medical case, and it can be seen that an alternative way of construing "respond[ing] to losses more than gains" is to say that the agent finds the losses involved in a choice situation more salient than the gains due to the way in which they are presented to her. So, referring back to the textile analogy, the losses are more "woven" into her decision-making than the gains. System 2 processes do not, however, help to illuminate the context of choice since they are "conscious, slow, controlled, deliberate, effortful, statistical, [and] suspicious" (Shleifer, 2012, p.3) and are thereby agent-based rather than being context-based.

It can be seen that in the negative account of the theory of BR, alterations in the context of choice act on the agent in a harmful way as a result of changes in salience to her of features of the choice situation which result in her making less than optimal decisions. The positive claim about BR is that this problem can be addressed through the agent drawing on strategies or heuristics to help her to arrive at the most rational decision that she can, and it is to this that I now turn. As with the negative claim of BR, there are features of the positive claim that help to shed light on the context of choice, in particular its reference to the environment, or choice situation.

---

[4] A further objection to Kahneman comes from Annas, who claims that if Kahneman's ideas are accepted into BR this opens it up to the criticism that it leaves no space for habituated virtuous behaviours that are neither System 1 nor System 2: "[v]irtue obviously isn't mindless habit, but equally it's not a matter of consciously controlling and directing yourself all the time, an annoying soundtrack to all your activity" (Annas 2015 p.3).

4.7.3 The Positive Claim in Bounded Rationality

In the first account of BR, Simon (1957) argues that it has two aspects. These are the agent's rationality, as discussed above, and the agent's environment. Furthermore, Simon claims that the environment has a positive role to play in human decision-making. The two parts of BR in Simon's account are introduced by an analogy with a pair of scissors: "[h]uman rational behavior […] is shaped by a scissors whose two blades are the structure of task environments and the computational capabilities of the actor" (Simon 1990 p.7). Newell and Simon further develop the environmental blade of the scissors in "Human Problem Solving" (1973). Here, they claim that rationality is adaptive, and consequently human behaviour "is determined by the demands of that task environment rather than by its own internal characteristics" (Newell and Simon, 1973, p.149). In this account, the demands of the task environment are clearly external to an agent but also appear to be woven into her decision–making since they may determine her behaviour. This feature of BR thereby shows the potential to help illuminate the context of choice.

Newell and Simon show how rationality can be shaped by the task environment using a puzzle in which letters have been substituted for numbers in an equation:

DONALD

+GERALD

ROBERT

The task environment in this puzzle is shaped such that some parts of it—i.e. D + D = T at the far right—are less cryptic. These parts are thereby easier to solve, and the agent will be most successful in solving the puzzle if she uses a heuristic that first focusses on them. A much less successful approach would be to randomly assign numbers to letters and then work through all the different permutations (of which there are factorial 9—362,880). Fortunately, in studies of this problem, most agents assign numbers to letters in an order suggesting that they are following the heuristic recommended above. So, the DONALD + GERALD problem shows that agents' rationality is both bound—they cannot solve the problem by random assignment—and also shaped by the task environment in favour of the "solve the less cryptic parts first" heuristic.

When an agent looks at the less cryptic parts of a puzzle before the other parts, she is finding the former parts more salient and thereby weaving these into her decision-making. Furthermore, it is these less cryptic parts of the puzzle that should be more salient to her in order for her to be the most efficient in solving the puzzle. This example from the literature on BR, then, provides evidence of a way in which an agent may perform well in a choice situation as a result of using an heuristic which renders certain aspects of it more salient.

Grüne-Yanoff, however, claims that "[i]n later papers […] Simon expressed doubts about the separate relevance of the environment" to BR (2007, p.552). In defending this claim he quotes a passage from Newell and Simon's "Human Problem Solving" (1973):

> "It is precisely when we begin to ask why the properly motivated subject does not behave in the manner predicted by the rational model that we recross the boundary again from a theory of the task environment to a psychological theory of human rationality. The explanation must lie inside the subject: in limits of his ability to determine what the optimal behavior is, or to execute it if he can determine it. [(Newell and Simon, 1973, pp.54,55)]"

In response to Grüne-Yanoff, we can say that the re-crossing he refers to, from the "theory of the task environment" to the "theory of human rationality", does not imply that there is no separate role for the environment in explaining the agent's behaviour in a choice situation: environment and rationality are both important and intersect with one another. So an agent with bounded rationality may be more susceptible to making errors in certain environments or contexts. Shortly following this passage in "Human Problem Solving" is another section in which Newell and Simon summarise the respective roles of rationality and environment in BR.

> "1. To the extent that the behaviour is precisely what is called for by the situation, it will give us information about the task environment. By observing the behaviour of a grandmaster over a chessboard, we gain information about the structure of the problem space associated with the game of chess.
>
> 2. To the extent that the behaviour departs from perfect rationality, we gain information about the psychology of the subject, about the nature of the internal mechanisms that are limiting his performance" (Newell and Simon, 1973, p.55).

This passage indicates that, for Newell and Simon, environment does have a role in BR. This role appears here to be a *positive* one in shaping rationality, as in the DONALD + GERALD experiment, rather than a negative one in explaining how rationality is bounded. So the way in which a Grandmaster behaves in a game of chess is dictated by the details of the game of chess, or in other words, facts about the relevant situation.

A further aspect of the positive claim in BR is defended in Simon's account of "satisficing" heuristic (Simon, 1957, p.241). The purpose behind Simon's account of satisficing is that he is trying to identify a type of rational behaviour "that is compatible with the access to information and the computational capacities that are actually possessed by organisms, including man, in the kinds of environments in which such organisms exist" (Simon, 1957, p.241). This type of rational behaviour, or satisficing, also illuminates the intersection between environment and rationality mentioned above, since in complex choice environments an agent with limited rational powers accepts what she feels is the best choice available at that point in time—she "satisfices". In such cases, the features of the choice environment that are salient to the agent are those that enable her to fulfil the requirements of satisficing, and these will shape her choice.

Simon illustrates his account using an example from a game of chess in which a player does not look for further possible moves after she has identified one that will lead to victory through forced check-mate. There is no need after identifying the forced-mate move for her to find an alternative sequence of moves that will result in conventional check-mate. A further example of satisficing is of agents who are trying to find their life partner and who, rather than embark on a comprehensive search of all possible candidates, stop searching after finding the first candidate who appears to satisfy their requirements. These two examples of satisficing are different in nature, and the difference brings out some of the problems faced by an agent who satisfices.

In the chess example of satisficing, the agent has a clear goal of winning the game, and an unequivocal means to this end; there is no doubt in chess that forced-mate will win the game. In contrast, in the life-partner example the agent may be in some doubt as to where to stop her search—she has to set an "aspiration level" (Simon, 1957, p.253) at which point her preferences are satisfied. Furthermore, the agent may adjust her aspiration level upwards if it proves easy to find suitable life-partners, or downwards if it proves difficult. One problem with setting an aspiration level is that it, too, involves the computational costs of its assessment against values held by the agent. Further analysis of the chess example shows that it is also prey to the same problem: the chess player will normally (until the later stages of the game) be unable to see a move that will lead to check-mate because of the complexity of the environment. As a result, she will also have to evaluate a range of different moves that fall short of achieving her goal of winning. So perhaps satisficing is workable only when the computational costs of

setting an aspiration level are lessened, e.g. in situations where there is less doubt that the means chosen will clearly achieve the end, as in the end-game check-mate case.

In the examples of satisficing, above, if an agent who wishes to make use of this heuristic should, first, find salient the difficulty of making an ideal choice that is in line with her preferences. Second, she will need to find salient features of the choice situation that will shape her decision-making to at least satisfy her preferences. Third, from what I have claimed above about computational costs, she should find salient the net costs in setting her aspiration level.

4.7.4 Conclusion

I have defended BR as a concept with both negative and positive claims. The negative claim explains how rationality is constrained, resulting in non-ideal choices, and the positive claim is that partial guidance—e.g. through heuristics such as satisficing—can be used by an agent with limited rational resources in a complex choice situation to make the best choices possible. The light shed by the theory of BR on the context of choice is that, first, there are features of a choice situation that should be salient to an agent because it is these features that should shape her decision-making. Second, the theory shows that the context of choice can be altered so that features of the choice situation are rendered more or less salient to the agent, in ways which can either be beneficial or harmful. I now turn to the second literature that may illuminate both the context of choice and potential harms to an agent who is offered additional options, that of adaptive preferences (AP).

4.8 Adaptive Preferences

I will argue that the theory of adaptive preferences (AP) sheds light on the context of choice because it provides further evidence that the context of choice can be altered in such a way that what is salient or not salient to an agent in a choice situation is also altered. The theory of AP includes accounts of choice situations in which changes or constraints in the context may exert an influence on an agent, namely the agent's preferences may change or form in order to cope with these changes or constraints. So, when the range or quality of options on offer is either altered or constrained, the agent adapts her preferences and in so doing what is salient or non-salient to her in the context of her decision-making is altered.

The AP literature focusses, first, on choice situations where it is assumed that the range or quality of options is altered in a way that is detrimental to an agent, i.e. reduced—as in cases of acquired disability or kidnap. So in paradigm cases of AP involving agents who become disabled or are kidnapped, prior to being disabled or kidnapped the agent would prefer not to be in that state. However, after the change in the agent's state, her preferences may adapt to the new state, and this adaptation may persist even if there is a prospect of treatment or release.

Second, the literature on AP describes people who are born into a set of constrained options, so the options themselves do not alter, but are continuously present. An agent in this type of case may respond not by seeking to pursue what she desires, but by forming adaptive preferences as a way of coping with the restraints. Agents who are born either with a disability or into a repressive social hierarchy, such as those in parts of India (Nussbaum, 2001), are examples of this type of case. I note here that cases such as these, in which an agent's range of options has always been restricted, may be evaluatively different from cases in which her options are restricted later in life (Berges, 2011). However, the agent in the former case still undergoes an adaptation of her preferences to her circumstances and that alters what is salient to her in her choice situation.

In the following sections I will give an account of the theory of AP, beginning with a brief account of preferences themselves, and then moving on to describe how preferences may be adapted and formed in response to a relevant context.

4.8.1 Preferences

The following example demonstrates the way in which preferences may be formed. Consider Paul, who is offered either an apple or a banana at a picnic and chooses to take the apple rather than the banana. In this situation, where it is assumed that there are no factors other than the fruit that may affect his choice, Paul is being asked to consider two alternative states in the world: one in which he eats an apple and one in which he eats a banana. When he considers these two options and freely chooses to eat the apple he is expressing a preference for the apple over the banana—he prefers the apple and chooses it because, for example, in his opinion apples taste better than bananas. Alternatively, Paul could have chosen a banana rather than an apple, thereby expressing a preference for bananas. A final possibility in this situation is that Paul is indifferent about apples and bananas; he believes that neither apples nor bananas taste better than

each other, so he does not have a preference for one over the other. If Paul is indifferent and hungry, and does not equally dislike the taste of both apples and bananas, it is likely that he will randomly choose one or the other (perhaps he could toss a coin to help him choose).

The offer at the picnic gives one instance where Paul can express his preference about apples and bananas, but Paul may often find himself in situations where he is offered the same options. If Paul is of the opinion that apples taste better than bananas, then *ceteris paribus* on each occasion that he is offered either an apple or a banana he will choose an apple—his preferences in this regard seem stable. So when a person has a preference—say for A rather than B—he is rank-ordering A and B according to a shared property they possess (in my example, taste—an important property by which food is assessed), which renders A better, or more valuable, to him than B. At some time in the past, Paul formed his preference for apples over bananas by tasting both of them and making a decision about which of them tasted best. Note, however, that if the only fruit that an agent—call her Paula—has tasted is apples, she cannot know whether she prefers them to bananas, which she has never tasted. So in a situation where Paula is offered apples and bananas for the first time she cannot have a formed preference for apples over and above bananas as a result of a direct comparison. However, on tasting apples she may have decided that they taste so nice that she will always prefer them to bananas on this basis and not on the basis of a comparison. Alternatively, she could choose a banana in this situation, and on finding that she now finds the taste of bananas to be better that of apples, she may form a preference for bananas over apples[5].

In this section I have established that an agent's preference for one option over another is a rank ordering of these options, normally after a process of comparison based on a shared property of the two options. An agent may however prefer one option to another, without a direct comparison, if her assessment of the first is a sufficiently positive one. I now turn to a type of case, known as "sour grapes", in which an agent may adapt a preference that she has previously formed.

---

[5] This example shows a *benefit* of being offered an additional option. Paula benefits from the additional option of bananas because she now has the option of choosing bananas which she then prefers to apples.

4.8.2 Adaptive Preferences: Sour Grapes

The "sour grapes" fable by Aesop (Elster, 1983) sheds light on the context of choice because it highlights one type of situation, namely where the range of feasible options available to an agent is changed and she adapts her preferences so that what is salient or non-salient to her in the context of choice is also altered. In the fable, a fox with a sweetness-based preference for grapes sees some of them hanging down from a vine. After some time and effort the fox realises that he is unable to reach the grapes; he turns away and claims no longer to want them because they are too sour. Before his struggle the fox had a preference for grapes based on their taste, and also believed that the grapes were within his set of feasible options. After the struggle it is clear to the fox that the grapes are not within his set of feasible options and his preference for them is altered so that he no longer wants them. This adaptive preference results in a change in what is salient to him in the context of choice—from "grapes" to "the grapes are sour".

The change in salience to the fox of features of the context of choice may have come about through two different processes. One view is that this process could be a non-conscious "drive to reduce the tension and frustration that one feels in having wants that one cannot possibly satisfy" (Elster, 1983, p.25). Bovens, however, suggests that the fox may merely have "engage[d] in an act of self-deception" and "construct[ed] a false belief about a stable preference for a nonfeasible alternative" (Bovens, 1992, p.60)[6]. The importance of this fable to an understanding of the context of choice, however, is that it shows how alterations in the range of feasible options available to an agent can result in changes to her preferences, thereby altering the context of choice since what she finds salient is also altered.

4.8.3 Adaptive Preferences in Continuing Restrictive Situations

In addition to accounts of how preferences may be adapted in response to changing circumstances, the AP literature also gives accounts of contexts in which adaptive preferences may be formed in response to unchosen restrictive situations. These examples are relevant to an account of the context of choice as they illustrate how, for an agent who has adapted her preferences to her situation (rather than to changes in her situation), the lack of subsequent choices is not a salient feature of this situation.

---

[6] Bruckner's account of adaptive preferences is similar to Bovens'. However, Bruckner argues that certain adaptive preferences are beneficial and worthy of pursuit, but only if the agent is able to "positively endorse" them (Bruckner 2009, p.317).

Nussbaum (2001) and Berges (2011) have written about preferences formed by women who live in certain oppressive contexts, and Barnes (2009) has written about the preferences of people with disabilities, which I turn to in the next section. One example used by Nussbaum is of women living in parts of India, such as Andhra Pradesh, who are malnourished and do not have a dependable supply of clean water, or who are subject to financial abuse. Some of these women may accept their inferior position in society and "acquiesce[..] in a discriminatory wage structure and a discriminatory system of family income sharing" (Nussbaum, 2001, p.69). Nussbaum refers to these women as having "entrenched preferences" (Nussbaum, 2001, p.69).

The preferences of these women are formed during their lives as children and young adults in a context where the norm is for women to have an inferior position in society. They are brought up without a concept of equality with men, and as a result of adapting their preferences to this situation of non-equality, the possibility of gender equality, through education for example, is not salient to them. They do not thereby form their preferences in the light of an awareness of their actual capabilities[7]. Berges draws on arguments from Wollstonecraft and argues that the cases described by Nussbaum are examples of "stunted growth of preferences". Referring to historical accounts she claims that "women's preferences were 'grown' in a corrupt system which provided them with a distorted frame of references in which to form their values" (Berges, 2011, p.74). The distorted frame of reference described by Berges is one in which alternative lifestyle options for the women in question are not salient to them. They are harmed as their context of choice has been altered in such a way that they are not able to conceptualise any alternative ways of life; i.e. these alternative ways of life are not salient to them.

In the example above, the agent seeks to satisfy her preferences in the face of her situation and accommodates to it. However, an agent whose adaptive preferences are formed in oppressive contexts may come to change her preferences if her situation is changed, for example through acquiring information that becomes salient when

---

[7] Nussbaum's capabilities approach is relevant for my purposes because it makes a claim about features of a choice situation that should be salient to an agent in order for her to benefit. The capabilities approach, broadly construed, is a concept based on acceptance of universal norms for well-being; and as such it comes into conflict with arguments that an agent's individual preferences should be the basis for public policy on well-being. Nussbaum's list of capabilities includes life, health, bodily integrity, senses, imagination and thought, emotions, practical reason, affiliation, play, and control over one's environment" (Nussbaum 2001, pp.87-88).

previously it was not. An example of this is of women who are shown videos of others attempting "something brave and new", such as working in health care, and who then "develop the confidence that they can do something new too" (Datania 1992, quoted in Nussbaum 2001, p.68). Datania's women have had their range of feasible options increased, and then make choices as a result of the changes to their context of choice being salient to them.

In sum, this section has established that an agent's preferences may be adapted in response to her situation, or changes in her situation, and as a consequence, what is salient or not salient to her in her choice situation is also altered, in ways that are either harmful or not harmful. I now turn to the second example of AP mentioned above, which is of untreatable disability.

4.8.4 Adaptive Preferences: Disability

In this section I claim that an analysis of cases of disability within the AP literature illuminates the context of choice by showing how agents with a disability shape their preferences in the light of their situation, and how this alters the salience to them of features of the choice situation.

People with untreatable disabilities, referred to in Barnes' paper (2009), do not have the option of living without their disability, and so form attitudes to their condition and make choices within this context. In this respect, they are similar to women in Nussbaum's and Berges' accounts who are not given the option of equal treatment with men. However, people with disabilities normally coexist with, and are thereby aware of, the life conditions of people who are not disabled—these facts are salient to them when making choices. They are able to engage in thought experiments about how life could be, were they not affected by their condition, which may alter their preferences and consequently their choices. One possible outcome of this, is that the agents decide that even if it were possible for them to be cured of their condition, they would still prefer to remain in their disabled state. In this case, the imagined option of being able-bodied also has an influence on the agents' preference, which in turn alters what is salient or non-salient to them in choice situations.

Consider, for example, Janet, who becomes deaf in early adulthood, and who ten years later is offered a new cure for deafness. In the intervening period she has adapted to life without hearing and is richly fulfilled in her life as part of the deaf community, so

refuses the cure. It could be argued that she has been harmed by adapting her preferences because she is turning down the prospect of being able to hear again. However, according to Barnes, agents who adapt their preferences are not harmed by this unless there has been some kind of social distortion such as "abuse of power relationships, exertion of dominance, forcible removal by one party of another party's resources or freedoms etc" (Barnes, 2009, p.13). She also suggests that people who claim that an agent has been harmed by adapting her preferences are question-begging—in this case whether it is a harm to be deaf. They assume, in this example, that Janet's preferences are adapted to the sub-optimal state of being deaf. But whether or not being deaf "is in fact sub-optimal is precisely the question that is up for debate" (Barnes, 2009, p.7), and people often claim that being deaf is not sub-optimal[8]. For my purposes, however, it is merely necessary for me to argue that it is the salience or otherwise to the agent of different features of the choice situation that influences her choices. So if Janet's positive experience of her new life within the deaf community is more salient to her than the possibility of hearing, then it is the former that will influence her preferences and thereby her choice regarding treatment.

In sum, the concept of AP includes accounts of contexts or changes in contexts within a choice situation that result in an important psychological process in the agent, namely adaptation of her preferences. As a result of features of, or alterations to, these contexts, and consequent alterations of preferences, what is salient or non-salient to the agent is also altered. This supports the idea that the context of choice can be altered in such a way that what is salient or non-salient to the chooser is altered, and in ways that may be harmful or beneficial to the agent. In addition, the cases cited by Nussbaum of oppressed women who are shown that there are alternative ways of living suggests that there may be features of a choice situation that perhaps *should* be salient but are not, due to them being absent from the context of choice. Furthermore, the agent is harmed when they are not salient. However, Barnes' deafness case defends the claim that it is question-begging to make assumptions about features in a choice situation that either should or should not be salient to an agent. These cases suggest, then, that there may be subjective and objective aspects of salience in the context of choice, a claim I defend by way of a medical analogy.

---

[8] However, Barnes too may be question-begging. It can only be concluded that the deaf person's AP are not harmful if you hold a preference-satisfaction view of happiness or flourishing, but that just begs the question for example against a view based on flourishing, such as the one that can be found in Aristotle.

4.9 A Medical Analogy: Context of Diagnosis and Context of Choice

In medical cases, a Doctor has to make a diagnosis. In doing so there are features of the Doctor's context—call it the "context of diagnosis"—that are salient to her and may thereby influence the diagnosis that she makes. Through identifying aspects of the Doctor's context of diagnosis, that either should or should not be salient to her, I aim by analogy, to illuminate the concept of the context of choice more generally. Specifically, I claim that both features that should and features that should not be salient to an agent are part of the context of choice. The context of diagnosis is helpful here because there are aspects of the medical context that are relevant, due to being important for the health of the patient, and health is a normative concept. By analogy, then, there are also important normative aspects of the context of choice.

Consider a Doctor, Meriam, who is assessing a patient to make a diagnosis. Certain features of the patient may be salient to Meriam in light of both the training and further experiences she has had in her practice. Features of the patient that are salient to her may include his current symptoms, past illnesses, family history of disease, consumption of alcohol and tobacco, occupation, and other aspects of his social situation. The features that are salient to Meriam in the context of diagnosis are woven into the process of her diagnostic reasoning and thereby influence her in the diagnosis she makes: in effect Meriam chooses a diagnosis. In the same way, features that are salient to an agent in a choice situation are part of the context of choice and are woven into the process of choosing.

If Meriam is a good Doctor, then there are aspects of the context that *should* influence her diagnosis. For example, a persistent cough with weight loss in a person who is a heavy smoker should suggest to a good Doctor a possible diagnosis of lung cancer and, indeed, medical students are taught that this cluster of symptoms may indicate this diagnosis. These symptoms, which are properly salient to a Doctor who has been well-trained, are woven together with her diagnostic reasoning. Each of the symptoms "tugs" in a different way on her diagnostic process in order to make the diagnosis the correct one. So, there are aspects of the diagnostic situation that should be salient to a good Doctor, and these are part of the context of diagnosis. Fortunately, in this case, the cough, weight loss and smoking habit are salient to Meriam.

There may, however, also be features of the patient that are not salient to Meriam, but ought to be salient to her. Perhaps the patient worked with asbestos, which is highly relevant to a possible diagnosis of lung cancer since it greatly increases the chance of this diagnosis. Medical students are taught that in cases of possible respiratory disease it is important to ask the patient about exposure to inhaled pathogens, e.g. asbestos in the work place. Since the patient's asbestos exposure should to be salient to Meriam, but is not, it is also part of the context of diagnosis. Features of a choice situation, then, that are not salient to an agent but which should be, are also part of the context of choice.

Next in this account I consider other features of the patient that should not be salient to Meriam, since they are not relevant to the patient's current illness, such as having gall stones (as far as I know, there is no link between gall stones and lung cancer). This latest aspect of the context of diagnosis, namely features that should not be salient to the Doctor, helps to identify a further aspect of the context of choice—features of it that should not be salient to an agent.

Consider again the patient with respiratory problems who is seeking a diagnosis and treatment from Meriam. In a variant of the case the patient's partner, with the patient's permission, attends the consultation. The partner is confident and stresses that the patient worked as a volunteer in Malaysia several months ago; she wants the Doctor to consider an unusual tropical cause for the illness. In this variant of our case, all the other features of the patient remain the same; the only changes are that additional features, namely the trip to Malaysia and the partner's fear about a tropical illness, have been added. Meriam should know that the original features of the case point to a diagnosis of lung cancer, and she should also briefly consider and then rule out the relevance of the visit to Malaysia. Both the temporal relation and the pattern of the symptoms are inconsistent with an infectious disease. However, as a result of what the partner has said, the Malaysia trip may be salient to Meriam and cause her to commence investigations for an infectious disease when she should not do so. Both the partner's emphasis on a feature of the patient's background, and the partner herself, are additional features of the case which have become salient to Meriam and have altered the context of diagnosis. In this case, features of the choice situation that should not be salient to the agent are part of the context of choice and their salience to her has consequently had a harmful influence on the choice she makes.

As I have described above, in making her diagnosis, the patient's overseas trip is fully salient to Meriam when it should not be, and she acts in the light of this. It can be seen through variations of the case, however, that the salience of the overseas trip could be *proportionate*. If, perhaps, the trip was more recent and some of the patient's symptoms suggested an infective cause, then it should not be totally disregarded—Meriam should find it salient. She should, however, find the trip salient in proportion to its relevance to the diagnosis, which in this case is less salient than the other symptoms that indicate lung cancer. The standard for assessing the proper relevance of the trip would be determined by medical facts and causal links as they relate to the patient's condition. It can be seen, then, that the degree of salience to an agent of different features of the context of choice may be proportionate, according to their relevance to the choice being made.

A final aspect of the context of choice that is illuminated by the medical analogy is the interaction with each other of features of the context in ways that *reduce* their salience to the Doctor. So, if the lighting is poor in the consultation room then this may influence Meriam to miss visible signs of an illness, such as the rash of dermatomyositis that may be associated with malignancy. This type of interaction is a sensory one: poor lighting influences Meriam to miss visible signs that are dependent for being seen on adequate lighting. So we now have two ways in which the context of diagnosis can influence Meriam. First, there are features of the context of diagnosis that can increase the salience to Meriam of aspects that should not be salient—the partner's emphasis on the trip to Malaysia—and second, there are features that can influence Meriam not to find salient aspects that should be so—poor illumination of the rash.

The medical analogy has shown that there is are normative aspects to the context of diagnosis. In other words, there are features of a diagnostic situation that should be salient to a good Doctor and either are or are not salient to her. There are also features of a diagnostic situation that should not salient to her and either are, or are not. Last, there are aspects of the context of diagnosis that may interact with other aspects of the context of diagnosis to either render features of the situation salient or non-salient when they should be, or salient or non-salient when they should not be. It is possible to identify standards for this normative aspect of the context of diagnosis: the various shoulds and should nots are set according to the goals of health and well-being and by professional norms of good practice which, in turn, are based on medical facts and causal links. The medical analogy has also shown how it may be possible to generalise

about what makes a good or bad context of diagnosis. A good context of diagnosis will be one in which features of the situation that should be salient are properly salient and *vice versa*, thereby enabling the Doctor to make the correct diagnosis.

It is now possible to apply the medical analogy to the concept of context of choice. The analogy suggests that there are normative aspects of the context of choice if it has features that either should be salient to the agent or should not be salient to her. However, this does not give us any reasons why the agent should find these features either salient or non-salient. In the medical case, health-care norms can be applied in order to identify which features either should or should not be salient, but there are not always clear equivalents for agents in more general choice situations, for example when buying food. One potential approach is to say that an agent should find certain features of the choice situation salient (or non-salient) if they should be salient (or non-salient) in order for her to make a rational choice. So, although there may be norms that indicate what should be salient to her, for example in relation to health and well-being, an agent who has a preference for chocolate should find the chocolate in a shop salient to her when she chooses what to buy. The corollary of this is that the same agent, who is also not interested in novel, vegan super-foods, should not find these salient.

The medical case suggests that the proportionality or degree of salience to an agent of features in the context of choice determines whether or not she is harmed when making her choice. Proportionality in the medical case was determined by medical facts and causal links, but in the general case it should be related to the requirements for an agent to make a rational choice. The proper degree of salience to an agent of a feature of a general choice situation, then, should be the degree of salience to her of relevant features in previous similar choice situations, i.e. ones of comparable gravity, in which the outcome of the choice was, in the agent's opinion, a rational one. To return to the medical analogy, the Doctor will have had training and work experiences that will influence the degree to which she finds salient certain aspects of the diagnostic situation. In the general case, the agent will have had life experiences that have formed the standards that are rational to her. So these life experiences can help to determine which features of the context of choice should or should not appear salient to the agent and the correct proportion or degree of their salience to her, in order for her to make a rational choice. Nussbaum describes the process of appreciation of salience thus:

> "even if [she] comes […] with good principles, the case does not present itself with labels written on it, indicating its salient features. To pick these out, [s]he must interpret it; and since often the relevant features emerge distinctly only through memory and projection of a more complicated kind, he will have to use imagination as well as perception" (Nussbaum, 1986, p.42).

So Nussbaum is arguing that perception, memory of similar situations, and imagination—to confirm that the situation is relevantly similar—all play a part in appreciating salience. Wiggins agrees that the agent "may require a high order of situational appreciation or as Aristotle would say perception (*aisthesis*)" (Wiggins, 1975, p.43). Situational appreciation in Wiggins equates to Nussbaum's perception and interpretation, where interpretation is also dependent on memory and imagination.

As we have seen in the medical analogy, adding additional options to a choice situation may increase their salience to an agent, which may in turn potentially alter the salience of other features of the context of choice, thereby increasing the chance of harm to her. In choice situations that the agent has not previously encountered, such as ones where a novel, additional option is offered, there may be increased scope for the agent to appreciate as salient this potentially harmful feature or option. This is, of course, relevant to most end-of-life situations, which are the focus of chapter six.

So, from a medical analogy I have argued that the context of choice consists in features of a choice situation that, first, are salient to the agent and either should or should not be salient to her, second, are not salient to the agent and either should or should not be salient to her, and third, interact with other aspects of the choice situation to alter their salience to the agent. Whether or not features of the choice situation should be salient or non-salient to the agent is determined by relevant facts about the choice situation and the requirements for making a rational choice. Furthermore, this requirement for a rational choice determines the proportionality or degree of salience that features of the choice situation should have for the agent. Last, the medical analogy demonstrates how there is the potential for an agent to be harmed in a choice situation if she finds salient some feature of the choice situation that she should not find salient or does not find salient a feature of the choice situation that she should find salient. The analogy also shows how it may be possible to generalise about what may make a good or bad context of choice. A good context of choice will be one that enables the agent to make a rational choice.

4.10 Harms that may Accrue from the Context of Choice

I now appeal to the notion of the context of choice to explain the way in which an additional option[9] may cause harm to an agent who is offered such an option. As I will argue in chapter six, this is relevant to the case of PAS because if patient choice is extended by offering the option of PAS then this offer may cause exactly this type of harm to some patients.

4.10.1 Harmful Influence of Changes to the Context of Choice

The context of choice is relevant to the potential harm an agent may experience when she is offered an additional option because this additional option constitutes a change in the context of choice which may become salient to her in a way that is harmful. First, she may perceive the new feature of the choice situation as salient when she should not find it salient—e.g. the relative valuation of the two pens in the decoy effect. Second, she may not perceive as salient the new feature of the choice situation when she should perceive it as salient—e.g. if the feature has been presented in a particular way. The alteration of what is salient within the context of choice may alter the choice the agent makes from one that is rational to one that is harmful, because it may no longer be dependent on her interests or based on an objective assessment of what is relevant to the choice being made. There are two related harmful effects on the choice the agent makes, resulting from changes to the context of choice: first, the agent may be harmed as her choice may not be the same as the one she would have made before the change in the context of choice exerted its influence. Second, she may be harmed because her choice may be one which she regrets after having made it. I will now turn to the particular harms that result to an agent when the context of choice is altered through her being offered an additional, novel option.

4.10.2 Harmful Influence of a Novel Option

An agent may have no reliable way of determining what value, and thereby salience, she should confer on an additional option that is novel to her if she has not previously

---

[9] An agent may alternatively be presented with additional information about existing options rather than an additional option. There is an overlap between these two types of case since an additional option also entails the agent being provided with additional information—information about the option itself. So, in cases where the agent is simply presented with additional information and not with an additional number of options, then the additional information may have an effect similar to the offer of an additional option. This is because new information, whether it arises in an additional option or in existing options, may affect the context of choice.

experienced it. Furthermore, when an additional option is offered, this draws attention to, and increases the salience of, certain features of the context of choice, including the novel option itself. How the agent makes her choice may also be influenced by other effects of the introduction of a novel option, such as those described above in the empirical studies of positioning, framing and decoys. Finally, the valuations of the additional option by other people, including the person who has made the offer, may also be amongst these features.

By way of example, consider Jane, who is offered a novel option. As I suggested in chapter one, the behaviours of other agents in the choice situation may be salient to her. These agents face the same options as Jane, including the novel option. If Jane witnesses one other agent making her choice in the same choice situation, this agent may potentially act as an exemplar, thereby influencing Jane's choice. If, on the other hand, Jane witnesses many other agents choosing a particular option, this may also exert an influence on her as a result of them presenting a "social norm" (Elster 1989, p.99).

The effect on an agent, then, of being offered an additional option, may be that she wrongly or disproportionately finds salient the additional options or features that comprise the context of choice. The result of this is that, through making her choice in response to an additional option, the agent may be harmed.

4.10.3 Configuring the Context of Choice

The various ways in which an agent potentially may be harmed as a result of being influenced by the context of choice suggests that there may be ways in which the context of choice can be configured to help agents to make a *good* choice. The context of choice could be set up with the agent's interests and well-being in mind, perhaps offering a reduced range of options in line with her preferences, rather than with an emphasis on her autonomy. In the latter case the emphasis would probably be on offering as many options to the agent as possible since, on one construal, agents are simply autonomous choosers, so maximising choice would be the most important value. The former case, on the other hand, could be construed as undermining the autonomy of the chooser; whilst nevertheless conferring a benefit or reduction in harm to the agent as a potential outcome.

Consider a medical example of configuring the context of choice to reduce the chance of harm to a patient. A Doctor who has a good knowledge of his patient and the way he

makes decisions, may present treatment options to him in a way that is in line with his interests. For example, the Doctor may know that the patient has a consistent aversion to surgical treatment options, and in the light of this she may avoid presenting to him an extensive range of highly invasive surgical options and data about their possible outcomes. Further examples of modifying the context of choice in order to protect people from harm can be found in the financial world where people may be presented with options, perhaps for pension schemes, that are in line with their interests.

4.11 Conclusion

This chapter has argued that a second aspect of choice situations (in addition to types of weak character)—namely the context of choice—may constitute a type of harm to an agent when she is offered an additional option. In order to argue this, first, the context of choice was delimited to features of the choice situation that are woven into an agent's choosing in virtue of their salience or otherwise to her in making a rational decision. Second, the literature on bounded rationality was analysed and found to support two arguments about the salience to an agent of aspects of a choice situation. These are that an agent's decision-making may be adversely influenced if she finds certain features of the choice situation salient when they should not be salient to her; and that there are features of a choice situation, identified by using heuristics, that should be salient to the agent in order for her to make the best decision. Third in this chapter, cases from the literature on adaptive preferences were drawn upon to establish that a choice situation can be altered in such a way that an agent may adapt her preferences and, as a result of this, the features of the situation that are salient or non-salient to her may also be altered. This supports the claim that the context of choice can be altered so that what is salient or non-salient to the chooser is altered in ways that may either be harmful or beneficial. Fourth, a medical analogy was used to defend the claim that the context of choice consists in features of the choice situation, first, that are salient to an agent and either should or should not be salient to her, and second, that are not salient to her but should be salient to her. The medical analogy also established that there is potential for an agent to be harmed in a choice situation when the context of choice is altered, for example through introducing an additional option. This is because the alteration may result in her finding salient a feature of the choice situation that she should not find salient, or not finding salient a feature of the choice situation that she should find salient.

I have now identified two key aspects of choice situations, namely types of weak character and adverse aspects of the context of choice, that may result in harm to agents when they are presented with an additional option. I am thereby in a position, in chapter five, to argue that these harms can be weighed against the benefits to some agents of being presented with the same additional option.

Chapter 5. Weighing Harms

5.1 Abstract

My purpose in this chapter is to defend an argument that some types of harm have the status of trumps with respect to other types of harm. That is to say that some harms, by their nature, should always take priority over harms of relevant other types. This is important to my thesis because I defend a claim in chapter six that the harm of wrongful death, should PAS be made permissible, should always take priority over the harm of unbearable suffering, should it not be made permissible.

I have argued in chapters one to four that an agent may potentially be harmed if she is offered an additional option, and that in addition to the types of harm identified by Dworkin (1988) and Velleman (2015), there are two further types of harm which may occur—that of weak character and that of context of choice. This argument does not rule out the possibility, which I do not defend, that a different agent may be harmed if she is *denied* the same additional option. The harm in the latter case would be that of being denied the potential benefits inherent in the additional option. So there can be cases where on the one hand there are potential harms to agents if an additional option is offered to them, and on the other hand there are potential harms to other agents if the same option is not offered to them. One approach to answering the question of whether or not an additional option should be offered, in a case where its result is harm to some and benefit to others, is to weigh these two harms against one another to see if one of them may outweigh the other.

One factor to consider in conflict cases such as these where the harms to two groups of people are being weighed, following Taurek (1977), is the respective numbers of people on the two sides. A further factor is the qualities of the harms on either side. So the harms on either side could, for example, be either equally severe or one could be more severe than the other. Both equally severe harms and those which differ in severity can also vary in their significance: some harms are trivial, but others are much more grave. I argue, however, that it is not possible to infer from arguments about features of these conflict cases that there are some harms that are trumping harms. Furthermore, it is not possible to infer from accounts of harm itself that there are some harms that are trumping harms. So in order to defend my claim that some harms are trumping ones, I turn to a parallel line of argument about needs from Wiggins (1998) and Megone (1992)

who claim that categorical needs, in virtue of being necessary for human survival, should always have priority over instrumental ones. Following this argument, I claim that categorical harms are trumping ones, and should always take priority over non-categorical harms.

5.2 Introduction

This chapter is a defence of the claim that some harms are trumps and should thereby always take priority over non-trumping harms. The first step in my argument is to claim that the literature about both weighing harms, e.g. Taurek (1977), and harms themselves, e.g. Feinberg (1987), does not allow us to infer that some harms are trumps. My next step is to follow a parallel line of argument about needs. Categorical needs, after Megone (1992) and Wiggins (1998), always take priority over instrumental ones. I will argue that harms that are parallel to categorical needs are trumping ones and should, by their nature, always take priority over other harms. Ronald Dworkin's argument about "rights as trumps" (1989) is another potential resource for my argument about trumping harms. However, despite being supportive of the idea that some harms may be trumping ones, Dworkin's argument, first, does not tell us what counts as a trumping harm, and second, is restricted to protecting rights in the face of a utilitarian regime, so it is not clearly applicable to the types of case that are my main focus.

I will argue first, then, that the question of whether the numbers should count in Taurek cases has not been finally resolved. If it had been finally resolved, in favour of Taurek's (1977) argument that the numbers do not count, then this would give us reason to think that the harm of death may have a special status—a status similar to the type I am looking for that should give it priority over other harms. Taurek cases are ones in which the claims to rescue of two differently-numbered groups of people in a life-threatening situation cannot both be met, and a rescuer thereby has to decide which group to rescue. According to Taurek, the status of the death of one person is sufficiently grave that it should be on a par with the deaths of a larger group of people. However, the debate over whether or not the numbers should count in these cases has not been resolved because there are problems on both sides of the argument. On the one hand, it is difficult to find a perspective from which it is possible to aggregate all the different claims to rescue in order both to weigh them and to prioritise the more numerous side for rescue. On the other hand, Taurek's suggested decision-procedure leads us to the implausible

suggestion that the rescuer should toss a fair coin, even in cases where one side is much more numerous than the other.

Kamm (2005) has argued, against Taurek, that the aggregation of claims to rescue is possible. Her argument is that it is possible to provide a perspective for aggregation based on Pareto-optimality and substitution of persons. If Kamm's argument succeeds then this would give support to a claim that the harm of death does not have a special status of the type that Taurek claims. However, following Lübbe (2008), I will argue that Kamm's argument fails because she herself is reluctant to capitalise on her substitution move.

Lang (2005) argues for a mixed approach to weighing the harms to two groups of people. His argument is that the numbers count, but fair treatment may sometimes be determined by tossing a fair coin. In particular, Lang uses the concepts of selection unfairness and outcome unfairness to explain how in most conflict cases the numbers do count. However, Lang's account does not deal with conflict cases that pit large groups of people against unequal but similarly numbered groups. In sum, Lang's argument is not useful for my purposes because it does not settle the question of whether some harms are trumping harms.

A second relevant issue, after the numbers of people facing *lethal* harms on the two sides, is how to decide in cases where different types of harm must be balanced. These types of case may be relevant to my thesis if they reveal that certain harms may be trumping ones. However, I defend a claim that existing arguments about the weighing of harms that are of a different type, and which thereby have a different significance, are limited in their applicability to the idea of harms as trumps. This is because the types of harm that clearly give way to more significant harms are "trivial" ones (Kamm, 2005), and there is often a need to balance different non-trivial harms. When more significant but different harms are balanced, Scanlon's "individualist restriction" (1998) is capable of offering guidance in some cases, but in others there may be a stalemate between significant (but different) claims to rescue in which the agents on both sides feel they are entitled to help. So this part of the literature on harms, also, does not clarify whether some harms may be trumping ones.

A further potential source for arguments about weighing harms, and in particular, harms as trumps, is the literature on harms themselves. A representative selection of the

literature on harms, however, does not allow us to infer that some harms are trumping ones. So I turn in the remainder of the chapter to an argument on needs that is parallel to an argument on harms. Categorical needs, after Megone (1992) and Wiggins (1998), always take priority over instrumental ones since they are grounded in basic facts about human survival. Instrumental needs, on the other hand, are elliptical for other needs. So I claim that some harms, such as wrongful death, that are parallel to categorical needs, are trumping ones and so should always take priority over other relevant harms.

5.3 Taurek Cases

My main claim in this section is that the arguments about Taurek (1977) cases have not been resolved, so they do not give us reason to think that some harms may be trumping harms, which is my overall interest in this chapter. Taurek has argued that in order to treat the two sides in a conflict case fairly, the relative numbers on each side should not count. The type of conflict case he considers is one where both sides are facing a lethal harm. In support of his claim, he sets up a rescue case involving five people who all need a drug to survive, and a rescuer who has a limited supply of the relevant drug. One of the people—"David"—needs all of the drug, and the other five people each need one fifth of the drug. He continues:

> [h]ere are six human beings. I can empathize with each of them. I would not like to see any of them die. But I cannot save everyone. Why not give each person an equal chance to survive? Perhaps I could flip a coin. Heads, I give my drug to these five. Tails, I give it to this one. In this way I give each of the six persons a fifty-fifty chance of surviving. Where such an option is open to me it would seem to best express my equal concern and respect for each person. Who among them could complain that I have done wrong? And on what grounds? (Taurek, 1977, p.303)

Taurek first argues that being required to count the numbers of people in order to make a decision is unlike merely attaching importance to the relative numbers of objects under threat of damage. In a case of deciding whether to rescue one object or five objects from a fire, where all six objects have an equal value, it makes sense to rescue the greater number "[b]ecause the five objects are together five times more valuable in my eyes than the one" (Taurek, 1977, p.306). In a rescue situation involving people, however, Taurek argues that it is the value to the people themselves of their continuing existence that is important rather than the objective value of each of the people: "it is the loss *to this person* that I focus on." Taurek continues by saying that the loss to an individual of his life is not a larger loss if it is accompanied by the loss of four other lives, and concludes by saying that "[f]ive individuals each losing his life does not add

up to anyone's experiencing a loss five times greater than the loss suffered by any one of the five" (Taurek, 1977, p.307).

So, for Taurek, a fair decision-making procedure in cases like this is not one that consists in an aggregation and balancing of claims, but one that gives each person in the rescue situation an equal chance of survival. If the rescuer tosses a fair coin then each person has a fifty percent chance of rescue. Taurek also argues that his method should apply even in cases which are more numerically imbalanced, such as one against fifty or more. So this line of argument suggests that in types of cases like this, the numbers should never count for the rescuer in her decision-making. Taurek thereby implies that lethal harms have a special status. If one person may potentially suffer a lethal harm then this should be given equal status in a decision procedure with more than one person who may potentially suffer a lethal harm.

5.3.1 Responses to Taurek: Kamm

Kamm argues against Taurek that the numbers should count in this type of conflict case. This is relevant to my overall argument since if the numbers should count then Taurek is wrong to afford a special status to a single person who may potentially suffer a lethal harm. Kamm uses her "Argument for Best Outcomes" to show that aggregation of claims on the two sides is possible:

> "(1) Using Pareto Optimality, we see that it is worse if both B and C die than if only B dies, even though it is not worse for B. That is B+C< B.
>
> (2) A world in which A dies and B survives is just as bad as a world in which B dies and A survives. This is true from an impartial point of view, even though the worlds are not equally good for A and B. That is, there is moral equivalence in the death of A or B.
>
> (3) Given (2), we can substitute A for B on the right side of the moral equation in (1) and get the result that it is worse if B and C die than if A dies. That is, if B+C<B and A=B, then B+C< A. Alternatively, we can substitute A for B on the left side of the moral equation in (1) and get that A+C< B" (Kamm, 2005, p.4)

Pareto-optimality is a state where "there is no alternative that is Pareto-wise better; that is, there is no alternative that everyone will regard as at least as good, and which at least one person will regard as better" (Blackburn, 2005, p.268). Using this principle, Kamm claims that moving from a world in which only B dies to a world in which B and C die is not Pareto-optimal since not everyone will regard this world as at least as good, and there is no one who will regard it as better. Lübbe asks from whose point of view this assessment of betterness is being made, and quotes Kamm who says that there is an

"impartial point of view" (Lübbe, 2008, p.73). In fact, Pareto-optimality implies an impartial point of view since it includes a stipulation that "everyone" evaluates the states of affairs under consideration. So, this first part of Kamm's argument already includes a numbers based conclusion that depends on an impartial point of view, namely the assumption that two deaths are worse than one. It can also be seen that it is not possible to construe this part of Kamm's argument as a rescue case. If we try to construe it in this way then B should not appear on both sides (see "(1)" above), and if it is possible to rescue both B and C rather than only B, then it is implausible to claim that anyone would want to rescue just B. The point established here is that Kamm merely assumes an impartial point of view for the purposes of her argument.

The next part of Kamm's argument is about the moral equivalence of two deaths: those of A and B. However, there are no conditions attached to this claim; there is no context, such as one where either A or B's survival is linked in some way to the survival of other people. The point of conflict cases is that the survival of the people involved is linked with the survival of other people, so Kamm has to show that, even if A's death is as bad as B's in some cases and *vice versa*, this equivalence can hold in a conflict case. As we will see below, Kamm appears reluctant to make use of the equivalence of A and B by substituting one for the other.

The last part of the "Argument for Best Outcomes" is the substitution of A for B in the Pareto-non-optimal (1): "That is, if B+C<B and A=B, then B+C< A" (Kamm, 2005, p.4). Lübbe focusses on what Kamm says about this substitution: "the conclusion [made by Kamm] is reached by substituting equivalents; but the respectful attitude forbids precisely that we regard individuals as substitutable equivalents" (Lübbe, 2008. p72). So according to Lübbe, Kamm appears reluctant to accept her own conclusion because of the lack of respect for A and B implied by treating them as mutually substitutable.

I have shown in the above that Kamm's argument that the numbers should count in conflict cases is dependent on there being both an impartial point of view from which to make assessments of betterness, and also on a substitution move that she herself is reluctant to utilise. Kamm merely asserts her assumption of an impartial perspective in the first part of her argument, and does not clearly defend it. So the key point established in this section is that Kamm has not refuted Taurek's argument (which is itself problematic) that the numbers do not count.

5.3.2 Responses to Taurek: Lang

Lang proposes an alternative, fairness-based approach to resolving conflict cases. If a coin is tossed in a conflict case, as *per* Taurek, then all the potential victims have an equal chance of being saved: fifty percent. So, coin tossing shows "selection fairness". However, an objection to selection fairness is that it "only assigns [to people] equal chances of being treated unequally, since only some of them will be saved, whereas others will not be. While chances of being saved with coin tossing are equally distributed among individuals, outcomes will be unequally distributed among them" (Lang, 2005, p.334). According to Lang, the outcome of a decision-making procedure is another source of fairness—"outcome fairness". Outcome unfairness is the unequal distribution of outcomes, mentioned in the quote immediately above. If the objective of the decision-procedure in a conflict case is to treat the potential victims in the fairest way, then both types of fairness should be satisfied. Any proposed decision-procedure should also be tested for fairness in a range of different cases. These should consist in cases with different numbers of people on the two sides.

The two principle decision-procedures I have been considering are coin-tossing and saving the greatest number. When these are assessed for either selection or outcome unfairness in different conflict cases they both have advantages and disadvantages. In a one-against-one case there is no option to save the greatest number and coin tossing is the only non-arbitrary way (i.e. not based, for example, on a whim) to choose who to save. Coin-tossing satisfies selection fairness in this case but not outcome fairness: "coin tossing means that actual harm is distributed in a way which cannot be traced to any inequality in the individuals' claims" (Lang, 2005, p.340). However, when the numbers on the two sides are unequal—as in one against many cases—coin-tossing requires "increasing indifference to actual outcomes" (Lang, 2005, p.341), whilst any potential criticism of the saving the greater number strategy in terms of selection unfairness is partially mitigated because the larger number of people who are actually saved have the compensation for missing out on selection fairness of actually being saved.

Lang admits that his analysis of decision-procedures in conflict cases cannot easily resolve those cases where the numbers are evenly matched and numerous e.g. in a case of one-thousand against one-thousand and one. This is because the one-thousand can complain about selection unfairness if the decision procedure is to save the greater number, and if a coin is tossed then there is outcome unfairness for an almost equally

large group. Since Lang is objecting to Taurek, at this point in my argument we are still left without an indication that some harms may have a special status with respect to other harms.

I have referred above to conflict cases where the harms to the individuals on either side are both equivalent and lethal[1]. I explained how Taurek argues that the numbers do not count because, according to him, there is no perspective outside that of each individual agent from which to aggregate the claims to rescue. Taurek's argument also leaves us with the unpalatable suggestion that we should toss an unbiased coin, even if there are very many more people on one side than the other. Kamm argues that aggregation is possible, but is reluctant to act on a fundamental part of her argument. Lang's account using selection and outcome unfairness helps in most conflict cases, but cannot help with imbalanced, many against many cases. So, we can tentatively conclude from these arguments that the numbers should perhaps count in some cases, such as where very many people are being balanced against a few, but coin-tossing may be appropriate in cases where the numbers of people are more equal. However, just what the decision-procedure should be in intermediate cases, where the numbers on the two sides are not either greatly, or narrowly, imbalanced is still contested. I have established in this section, then, that arguments about conflict cases do not lead us to infer that some harms may have a special status that makes them trumping harms.

5.4 Do the Numbers Count when the Harms are of a Different Type?

Taurek has argued that in conflict cases the relative numbers of people with competing claims to be rescued from dying should not count when the rescuer is deciding who to save. In tentative reply, it has been argued that it may sometimes be possible to aggregate claims between groups of people to justify saving the larger number. However, we have seen that whether or not the numbers should count in this type of case has not been decisively resolved. One important feature of these cases is that the harms involved (e.g. loss of life) on the two sides are identical. This makes it possible, if all other things are equal, to balance cases where the numbers are the same on the two sides, and to toss a fair coin as a method for selecting who to rescue. However, there

[1] There are also cases where the harms on both sides are equivalent but non-lethal (unlike typical Taurek cases). The arguments about equivalent lethal harms also apply in this type of case, since there is nothing about harms of different severity in and of themselves that make them differ in respect of whether or not they can be aggregated.

will also be, of course, cases where the harms on the two sides are not of the same type[2]. The main claim I make in respect of arguments about weighing harms in cases like these is that they, too, do not allow us to infer that some harms are trumping ones.

One approach to conflict cases where harms of different types are being balanced is to take an utilitarian line. An example of this in healthcare, is where the harm of suffering a painful arthritic hip, which could be cured by a joint replacement, may be balanced against the harm of suffering end-stage kidney failure requiring dialysis. Without dialysis, end-stage kidney failure inevitably leads to death within days or weeks, but an arthritic hip that is not replaced results in pain and disability, and not in death. It could be argued in this case that it is possible to add up a very large number of lesser claims for help, by the sufferers of arthritis, and then say that this is decisive. Others may respond, however, that no matter what the number of hip sufferers is, an illness that results in death should always take precedence[3].

An alternative approach to the utilitarian one is to argue that a lower number of greater harms may sometimes outweigh a higher number of lesser harms provided the agents with the lesser harms do not have a reasonable objection to being harmed. Conversely, if the agents with the lesser harms do have a reasonable objection to being harmed then their interests should be given serious consideration by the rescuer when she is deciding how to act (Scanlon, 1998). Scanlon (1998) argues that a larger number of less severe harms should only count when the agents who stand to suffer them would not willingly undergo the harm, despite being aware that should they do so, they would spare the smaller number of agents from a more severe harm. Scanlon's solution, however, cannot resolve a dispute between agents on opposing sides who both feel they have a legitimate claim.

Another approach, from Kamm (1993), applicable to otherwise balanced cases but where a much more minor harm is added to one side, is to discount the remedy for the lesser harm as an "irrelevant utility". I will argue, following Kamm (1993), that, despite

---

[2] If Taurek's argument about the numbers not counting in rescue cases where the potential harms are lethal ones had gone through, then this may suggest that in cases where the harms are not equivalent but in which the harm *on one side* is a lethal one, that the numbers should not count in these cases too and the side facing the lethal harm should always be rescued.

[3] One approach to this problem comes from health economics. The concept of Quality Adjusted Life Years (QUALYs) is grounded in patients' assessments of their needs *e.g.* via indexes of distress and disability, and in the duration of time over which the benefits of any proposed treatment are experienced. Though QUALYS purport to judge between claims for scarce resources they may fail to properly recognise some grave healthcare needs which should take precedence (Lockwood 1988).

their number, some lesser harms (or utilities) are irrelevant because they are not of the same order as the harms (or utilities) with which they are competing, either objectively or subjectively. Even when numerous, these harms should not be allowed to compete with greater harms because they are trivial.

Before I turn to Scanlon and Kamm in more detail, I outline a brief account of harm itself, and use a case from Hope (2004) to bring out the key features in conflict cases where the harms differ.

5.4.1 A Preliminary Account of Harm

A comprehensive account of harm is not necessary for my purposes here, since in this section we need only understand how harms may differ in severity. So I will work with an account which merely claims that a person is harmed at t_1 if they are worse off at some later time, t_2, than they were before t_1 (Meyer, 2016). Feinberg (1987) accounts for harm to a person by saying that she is harmed by suffering a setback to her interests; these could, for example, be "welfare interests" in the functioning of her body, or "ulterior interests" in some goal or aspiration. So, combining these two accounts, a person is harmed at t_1 if she suffers a setback to either her welfare or ulterior interests, such that either her body functions less well or she is, for example, unable to socialise with her friends at t_2. A case from Hope (2004, pp.26-41), illustrates difficulties that arise in weighing different welfare interests when some harms are more severe than others. In this case, these are on the one hand, an interest in being prescribed cholesterol lowering tablets and, on the other, an interest in renal dialysis. It is assumed, first, that the tablets confer a ten percent probability of reducing the future chance of atheroma, or artery-hardening related disease, such as a heart attack or a stroke. Secondly, it is assumed that there is not an organ available for transplantation and that, without an organ or dialysis, the latter patients will die from kidney failure within one month. For the purposes of this case it is also assumed that it is not possible to save both sets of people—resources are scarce and funds must only be spent on cholesterol tablets *or* dialysis.

In Hope's atheroma against dialysis case, the patient who does not receive the tablet suffers a setback in her welfare interests because she has an increased chance of suffering a stroke or heart attack, which would not normally be fatal but would affect the functioning of her body. However, the patient with kidney failure is certain to die without dialysis and death represents the most serious setback to both welfare and

ulterior interest. The harms in these two cases are clearly very different in severity, with relevance for my overall argument.

5.4.2 A Clinical Case of Balancing Different Harms

Having established that harms can differ in severity, I now situate some examples in conflict cases where the numbers are either balanced or imbalanced, and analyse arguments about how the claims of the agents on the two sides should be weighed. In doing so, I will be able to defend a claim that arguments about this type of case, still do not allow us to infer that some harms have a special status that makes them trumping ones.

In the clinical example above, end-stage kidney failure is much less common than atheromatous disease, so it is appropriate to consider this case as an example of a smaller number of people, here with end-stage kidney failure, competing for scarce resources with a larger number of people, who in this case need cholesterol lowering tablets. One potential approach to deciding who to help, or rescue, in this case is to aggregate the larger number of claims from the atheroma patients in order to outweigh the much smaller number of people who need dialysis. If the decision-maker agrees with an aggregation strategy and believes that there are sufficiently numerous claims for treatment from the atheroma patients to collectively respect them, then the dialysis patients will suffer being condemned to death. Conversely, if the dialysis patients are treated, then the patients with atheroma will suffer a potential setback to their interests in avoiding a heart attack or stroke (and they may also point out their superior number). The situation, then, is a stalemate.

5.4.3 Scanlon's Individualist Restriction

One potential response suggested by Scanlon to the stalemate above uses his "individualist restriction". The individualist restriction says that "an act is wrong when and because it is ruled out by principles that no one could reasonably reject from individual standpoints" (Hirose, 2007, p.278). The individualist restriction could be helpful to the atheroma patients; they could claim that the act of treating the dialysis patients is wrong because from their individual standpoints a principle that says they should be treated in preference to the dialysis patients cannot be reasonably rejected. The detailed content of the principle applying to this case is not explicit, and also appears to be question-begging about what is "reasonable", but it could refer to the

relative harms and numbers of patients on the two sides. Hope, however, expects that the atheroma patients may acknowledge the seriousness of the claims of the dialysis patients and then step aside in their favour. Alternatively, a health provider could reason as follows: "without treatment the chances are that the person [with atheroma] will not have a heart attack and die. By refusing the treatment we are not condemning him to death as we are the person who needs renal dialysis" (Hope, 2004, p.40). So perhaps we can reasonably reject a principle that says we should treat the atheroma patients. If, however, the more numerous patients instead faced a more significant harm that is almost on a par with death from renal failure, e.g. death from cancer within six months, then the individualist principle may rule out a preference to treat a fewer number of patients needing dialysis. So Scanlon's "individualist restriction" cannot help in cases where the harms are different to each other but more equivalent in severity. Furthermore, Scanlon's argument does not point to harms that may be trumping ones since an agent who is in competition with a putative trumping harm may just make a reasonable objection to a principle that says her competitor should be treated.

5.4.4 Kamm and Irrelevant Utilities

There may also be cases in which the more numerous group faces a lesser harm than atheroma, such as suffering a sore throat or a cold. The welfare interests of the patients in this case suffer a very minor setback that may be restricted merely to runny noses and a mildly uncomfortable change in quality of voice. Kamm has termed remedies for these minor harms "irrelevant utilities" (Kamm, 1993, p.146). In this section I will defend a claim that the concept of irrelevant utilities is not helpful in identifying harms that may be trumping ones. Since irrelevant utilities are trivial, they are unlikely to have any applicability to types of cases in which the harms on both sides are significant ones, and which are my main focus. Furthermore, Kamm's claim is about harms that should never take priority rather than harms that should always take priority.

Kamm introduces the idea of irrelevant utilities in a case where a threat is being redirected "away from five people in the direction of either Joe *or* Jim". After saying that a random decision procedure should be used to decide between Joe and Jim she elaborates the case by adding a patch of beautiful flowers beside Jim but not beside Joe. Now, if the threat is redirected towards Jim then the flowers will be destroyed and the potential pleasure that they may bring to "a great many people" will be lost. A further case describes a life-saving medicine that can be given to either Jim or Joe, "but only if

it is given to Jim will there be enough left over to cure a sore throat that Nancy would otherwise have had for a week (Kamm, 1993, p.146). Kamm defends a claim that the flowers and sore throat are irrelevant utilities by combining the subjective and objective standpoints in the case: her term for this combination is "sobjectivity" (Kamm, 1993, p.154).

Kamm argues that an objective aspect of the sobjective view is that the interests at stake between the two men do not differ too widely, and this is plausible: Jim and Joe both stand to lose their lives and neither the flower bed nor sore throat cure are comparable factors. The subjective aspect of sobjectivity is that people are "matched against each other by assessing size of expected loss to each rather than by aggregating just any set of losses, and (perhaps) that matching requires only certain losses to be approximately equal" (Kamm, 1993, p.154). This is also plausible since a self-interested agent will baulk at aggregation, but may also be sufficiently concerned about the other parties that she wants matched interests to be roughly equal or on a par. Furthermore, subjectively each agent is principally interested in his own survival and not that of his pair. A final objective component of sobjectivity is that Jim is seen as taking the same attitude to Joe as Joe takes to Jim.

If sobjectivity is applied to the sore throat case, in which there is only enough medicine left over to treat Nancy's sore throat should Jim, but not Joe, be treated, Joe appears to compete against Nancy's sore throat. The objective aspect of sobjectivity rules this out because the harms are not nearly equivalent or on a par. Kamm next claims that, subjectively, if either of Jim or Joe die then their experience is the same as if no one has survived; in other words, if neither Jim nor Joe survive then each of their experiences is the same as if it had only been them that died. If the sore throat cure (or the flowers) is then thrown in as a way to "alter this arrangement of a 50% chance for Joe and to determine that Jim shall be saved", then it seems that the sore throat cure is standing "as a contestant alone against him" (Kamm, 1993, p.155). Kamm's latter subjective claim is less plausible since it depends on a very pure sense of subjectivity in which an agent is unable to appreciate that his opponent will survive if he dies. Kamm admits in a footnote that it "is an exaggeration to say he cares nothing at all for the survival of others, but it is a useful exaggeration for this argument" (Kamm, 1993, p.164). However, this still leaves the objective claim that people should not compete against sore throat curing for their lives. But perhaps a subjective perspective is enough to rule the sore throat cure out of the balance. An agent, such as Nancy, may subjectively recognise that

her claim to treatment for her throat should not weigh in the balance, just as she may decide that if she is at risk of a heart attack or stroke without tablets then this need should not take precedence over another's need for dialysis treatment.

In the last two sections I have defended a claim that arguments from Scanlon and Kamm about conflict cases where the harms are of a different severity, do not help to identify certain harms as trumping ones. From Scanlon, an agent may reasonably reject a principle that guides the rescuer to aid another person who is in competition with her as, in this type of case, it may be assumed that the harm to the agent is likely to be more severe than the harm to the other person. However, if an agent stands to suffer a much less significant harm, then it is less reasonable in this case that she should be permitted to reject a principle that supports helping the agent who is facing the more significant harm. So even in cases where there are many more numerous claims for help, but for a much less significant harm than the competing one, then the less numerous but more significant harm should take priority. This latter case, however, does not amount to one in which the more significant harm is a trumping one, since the more significant harm does not invariably take priority over the less significant harm. Whether or not the one harm takes priority over the other will depend on what is seen as "reasonable".

Kamm's account argues about cases where trivial harms give way to more significant competing ones, so this limits the applicability of this decision-procedure to cases where one of the harms is trivial. In other words, her argument will not be applicable in the types of case that are my main focus and in which the harms on both sides are significant ones. Furthermore, Kamm's claim is about harms that should never take priority rather than harms that should always take priority. However, the notion that harms can differ from one another in severity may give us reason to think that the gravest harms may have some quality that should always give them precedence over lesser ones—that they may be trumping ones.

5.5 Do Some Harms Trump Other Harms?

One approach to solving the problems considered above, in which two groups of people are facing distinct and conflicting harms, and where only one of the sets of harms can be avoided, is to decide whether one of them trumps the other. I will defend a claim in the remainder of this chapter that some harms are trumping harms. If a harm is a trumping harm, then relief from it will always take precedence over relief from a lesser,

non-trumping harm despite the number of non-trumping harms. This is analogous to a trumping suit in card games always defeating a non-trumping suit, whatever the values of the respective cards within their suits. There are some practices in healthcare that can be seen to conform to this notion. Patients are normally triaged in emergency departments according to the priority assigned to their care. If a patient with a torn knee ligament and a patient with a ruptured abdominal aortic aneurism both arrived in hospital at the same time, and it was felt that the vascular patient's life could be saved, then she would always take priority over the knee patient. Furthermore, this would still be the case even if there were five, eleven or twenty-two knee patients waiting for treatment. In this case, then, the harm to the aneurism patient has the appearance of a trumping harm.

In order to investigate the claim that some harms may trump others, I will suggest three resources that may shine a light on the matter. First, there are accounts of harm itself. I will argue that different explanations of the concept from Feinberg (1987), Hanser (2008), Thomson (2011), Shiffrin (2012) and Kahane and Savulescu (2012) are indicative that some harms may be prioritised, but do not aim to settle the question of whether some harms may trump others. Feinberg (1987), however, offers a taxonomy of harms which may allow ranking and choosing between harms in balancing cases; he categorises harms as being related to interests—in particular vital, extensive, or morally weighty interests. Using this taxonomy, he claims that interests that are more vital, extensive, or morally weighty should take priority over interests that are respectively less vital, extensive or morally weighty. Prioritising in this way suggests which harms trump other harms. However Feinberg's taxonomy, despite indicating the notion of priority, does not set limits on this priority; it does not, for example, claim that there are harms that should always take priority over other ones, which is necessary for them to be regarded as trumping harms. In fact, Feinberg's taxonomy only allows for the prioritisation of harms within each category of harms, and not between them.

An alternative approach is needed to defend the argument that some harms are trumping harms. This second approach will draw on the concept of needs and, in particular, takes advantage of a relevant distinction between two kinds of need. Needs, after Wiggins (1998) and Megone (1992), can be either categorical or instrumental. Categorical needs, by their nature, are fundamental to continuing existence, so they should take precedence over instrumental ones (Megone, 1992). It can be argued that certain harms (such as the harm of wrongful death), are parallel to categorical needs, so should also take

precedence over lesser ones. This gives us reason to think that some harms should trump others.

The third resource that may shed light on the existence of trumping harms is an essay from Dworkin, R. (1989) in which he defends a claim that rights are trumps. Since, in one view, rights are grounded in interests[4], and are thereby related to harms, Dworkin's argument may also give support to the claim that some needs can trump others. However, Dworkin's claim about rights as trumps is limited to certain freedoms in a utilitarian state, and so lacks the scope necessary for my purposes. Furthermore, Dworkin does not fully explain the concept of trumping in his account; we cannot, then, make use of his argument to identify trumping harms.

5.6 Representative Accounts of Harm

A first resource that may be useful for defending the claim that some harms trump others is accounts of harm itself. Many of these accounts indicate that some harms are more serious than others. However, I will argue that these accounts do not aim to fully rank harms and thereby do not provide reason for thinking that some harms may trump others.

5.6.1 Feinberg

A suitable starting point is one of the more prominent modern accounts of harm, which is from Feinberg (1987). To develop his account, Feinberg narrows his focus from more general cognates of harm such as "damage", e.g. of a rock, to a normative account of harm, which he first conceives of "as thwarting, setting back, or defeating of an interest" (Feinberg, 1987, p.34). Feinberg compares having an interest in something to having a stake in it. If, for example, someone has a stake in a company, she wants it to perform well. The company is harmed if the stake is thwarted and the company fails to do well. Similarly, a person is harmed as a result of her interest in her well-being being thwarted[5]. There are two senses of interest in Feinberg's account: first, interests "as a miscellaneous collection, consist[ing] of all those things in which one has a stake", and

[4] A competing view to the interest account of rights, which argues that rights exist to further the right-holder's interests, is that they are instead based in the agent's will, which gives the agent "control over another's duty" (Wenar, 2015). However, my focus here is on the interest-based view.

[5] Note here that Feinberg's account links interests to desires. A problem for this account is that people can want what is not in their interests.

second, “interest or self-interest, [that consist] in the harmonious advancement of all one's interests in the plural" (Feinberg 1987, p.34). To illustrate this, consider Ann, who has an interest both in eating at least one good meal each day and in dressing warmly, as she needs to eat, lives in a cold climate and has to go out to work and to buy provisions for her family. These two interests in the first sense are coordinated and, alongside others, help Ann to advance her overall interest, which could perhaps consist in being a good parent. Ann would be harmed if her interests in food or warmth were thwarted and as a result she were unable to support her family. So far, according to Feinberg’s account, we can see that Ann has an interest in being satiated and warm, and is harmed if she is hungry and cold. The differences between being satiated and hungry, and between being warm and cold, suggest a key and defining comparative feature of Feinberg’s account which I outline below. This feature is his test for harms.

Feinberg’s test for harms depends on whether or not a person has had her interest(s) set back and asks “whether that interest is in a worse condition than it would otherwise have been in had the invasion [by self or others] not occurred at all" (1987, p.34). So, the account compares the outcomes in two of the closest possible worlds: one in which the person’s interests are set back by being thwarted, and the other in which there is no thwarting; it is termed the counterfactual comparative account of harm[6]. So, considering Ann again, if one day she is unable to eat a good meal, or she is unable to dress warmly before going out in the cold, perhaps because her income is insufficient, then her interests in both these things has been set back. Furthermore, she is harmed as a result of her interests in eating and being able to dress warmly being set back. In addition, her overall self-interest in being a good parent may also be set back since she is unable to work or to buy food for her children; she is thereby also harmed in this overall sense. If in the closest possible world to the world where her interests have been thwarted, however, she is able to feed and clothe herself, and to provide for her children, then by comparison she is unharmed in this counterfactual world.

Feinberg’s interest-based counterfactual comparative account of harm has faced objections, and it is important to analyse these objections and the alternative accounts of harm they generate because doing so may illuminate the question of whether some

---

[6] I am not seeking here to defend a possible worlds view, but aim merely to introduce a comparison between mutually exclusive outcomes that is made possible by a thought experiment where all the outcomes are possible.

harms may trump others. Some of these accounts do include the idea that harms may be ranked, however, I claim that none of them sets out to defend the idea of trumping harms.

One of the objections to Feinberg's account is from Hanser (2008, p.429), and refers to a case of "preventative harming". Hanser claims that the counterfactual comparative account of harm fails to differentiate between closely related but different types of harm. In a series of similar cases, Hanser describes a person who is harmed by losing her sight, but who can also benefit if she is prevented from losing her sight. In a further complication of the case, she can be prevented from being prevented from losing her sight. Hanser argues that the prevention of prevention case looks the same through the lens of Feinberg's counterfactual comparative account as the uncomplicated loss of sight case, but using an adequate account of harm it should look different. Feinberg's test of harm will just compare the outcome in the two cases—loss of sight—with the counterfactual case in which the person does not lose her sight. So Feinberg's account of harm does not draw out the difference in the two cases.

A further problem with Feinberg's counterfactual comparative account is that it fails to identify harms that may be different in severity and thereby open to prioritisation. In a footnote, Hanser does recognise that some harms may be more serious in kind than others, but he says this is not revealed if "a harm's seriousness is a function solely of the difference between the subject's actual level of well-being and the level of well-being he would otherwise have enjoyed" (Hanser, 2008, p.429). So the counterfactual comparative account of harm in and of itself, by not revealing the relative seriousness of harms, does not help to answer the question of whether or not some more serious harms may trump others. Feinberg may legitimately reply to this objection by saying that his account is not intended to answer that question.

As I have said, objections to the counterfactual comparative account have generated their own accounts of harm, which include Hanser's event based (2008), Thomson's (2011) state based, Shiffrin's will based (2012), and Kahane and Savulescu's (2012) norm based ones. These accounts, however, do not settle the question whether some harms may trump others. They merely vie with each other to give the most coherent explanation both of our intuitions about harming and whether or not a person has been harmed; they do not aim to systematically stratify or weigh harms. I will briefly analyse these accounts to show how none of them develops a taxonomy by which harms can be

given a weight, and thereby how none point to certain harms trumping others. Despite not aiming to fully develop a complete taxonomy of harms that enables their ranking, it will be apparent, however, that some of the accounts describe selected harms that take priority over others.

5.6.2 Hanser

Hanser's (2008) event based account attempts in part to overcome the problem for the counterfactual comparative account that following a person's death there is no subject to suffer its harm[7]. If there is no subject to suffer the harm of death then there is no state that the subject is in that can be compared with the state she would have been in had she not died. Furthermore, it is not possible to say at what time the subject who has died is suffering the harm of death. Hanser's solution is to ground his account of harm in events that result in the loss of a basic good, which people thereby have an interest in—in this case, the basic good of life itself. Hanser has no need to appeal to any harmful state of being dead: he says that the event of dying is itself harmful. And since the event that leads to death can be clearly timed, this is when the subject suffers the harm of death[8]. Hanser's event based account, however, merely explains whether or not a person has been harmed, such as by dying, and does not seek to explain how harms may be compared. People are harmed by the event that results in them losing a basic good, but he does not seek to rank the basic goods themselves.

5.6.3 Thomson

Thomson's is a state based account of harm (2011), and consequently has difficulty explaining the harm that is relevant to my overall purpose, which is the harm of death. After all, who is there who can experience the harmful state of death? Hanser accuses Thomson's account of the harm of death of being *ad hoc* since she says that "the harming one does in killing […] is a unique kind of harming for which the usual requirements for harming do not hold". She continues: "[a]ll other harmings […] cause the victim's life to be either non-comparatively or comparatively bad in a way. But when you harm a young and thriving person by killing him, you don't cause your

---

[7] In his Principle Doctrines, Epicurus (341-270 BCE) also recognised the problem in accounting for the harm of death that there is no subject who is subject to that harm. He says: "[d]eath [...] the most awful of evils, is nothing to us, seeing that, when we are, death is not come, and, when death is come, we are not" (Epicurus, quoted in Konstan 2016).

[8] One possible reply to Hanser is that there may be things that an agent has an interest in even after she has died.

victim's life to be bad in some way, you simply end it, which is a very different affair." Thomson's response to Hanser's criticism is to argue from a case of injury accrued during a rescue from a life-threatening fire. She claims that the person so rescued is better off being alive but with a broken arm than being "unbroken-armed and dead" (Thomson, 2011, p.455). Her response, however, by appealing to the converse state of being alive and broken-armed, allows her to make a claim about a person who, in virtue of being alive, is able to experience her state. She has not thereby directly confronted the problem of the absence of a subject after death.

Both Hanser and Thomson are arguing about how best to determine whether or not a person has been harmed (an argument that I do not need to settle), and neither of them attempts a formal categorisation of harms that may help in weighing them or determining which, if any, may be trumping ones. However, in Thomson's fire rescue case there is a claim that the harm of having your arm broken is less than the harm of death: it is of overall benefit to a person to have her arm broken if, as a result, she is rescued from a lethal situation. If, however, the person was certain to have her neck broken whilst being rescued from a fire, or alternatively was certain to be put into a vegetative state, then the rescuer may hesitate to proceed with the rescue because it would be less clear that the person would benefit from being rescued under these conditions. So there are some pointers in Thomson's account to harms of different levels of severity that may be useful in defending an argument that some harms may trump others.

5.6.4 Shiffrin

Next, Shiffrin's will based account of harm, too, does not attempt to formally categorise harms of different levels of severity such that some of them trump others. However, as with Thomson, Shiffrin's account gives an indication that some harms should have priority over others. Her account does this in two ways. First, in her criticism of the counterfactual comparative account, the case she uses contains two different harms that can be prioritised: "suppose that Jones shoots Black in the leg, but if he had not, Smith would have shot Black dead. On the counterfactual account, Black has not been harmed by Jones's shot. But it seems to me that Black's wounded leg and aching sores represent harms to her, although Smith's shot would have deposited her in far worse circumstances" (Shiffrin, 2012, p.368). So, being shot in the leg is harmful on Shiffrin's account (but not Feinberg's), but not as harmful as dying. Second, different harms in

Shiffrin's account can be subjectively prioritised by the agent herself. Shiffrin's account permits this because it is based on the will of the agent. She seeks to link autonomy and harm, and her motivation for doing so is the view, after Mill, that autonomy rights act as a general constraint on what we can do to people in order to promote their welfare or the welfare of the community. Furthermore, Shiffrin continues, "it is often impermissible for a third party to impose a harm to secure an overall benefit for a nonconsenting agent but permissible and even reasonable for that agent to make the same decision for herself" (Shiffrin, 2012, p.376). So Shiffrin attempts to reconceptualise harm and autonomy rights to "reveal greater connections between them" (Shiffrin, 2012, p.379). This reconceptualisation allows her to explain harm as a 'frustration or impediment of the will or of the ability to exert and effect one's will" (Shiffrin, 2012, p.383). This account allows that the agent prioritise different harms, depending on the degree to which her will is frustrated, impeded or unable to exert and effect itself. However, comparing harms between different people is problematic with a subjective will based account; things that are harmful for one person by frustrating her will may not be harmful by another's lights[9].

5.6.5 Kahane and Savulescu

Finally, I will argue that Kahane and Savulescu's account of harm does not clearly prioritise different types of harm in a way that suggests that some may trump others. Their account criticises the counterfactual comparative account on the grounds that a person is deemed harmed if she is in a state in which her Intelligence Quotient (IQ) is less than one-hundred and sixty, if she lacks great artistic ability or lives for less than one-hundred and thirty years (Kahane and Savulescu, 2012, p.319). So, to take the first of these listed states, if a person's IQ is one-hundred and thirty then according to a counterfactual comparison with a state of having an IQ of one-hundred and sixty, she is harmed. However, in contradistinction to the first list of states, Kahane and Savulescu claim that a second list consisting in being "severely intellectually impaired, paraplegic, blind, or to die in one's 20s is to suffer, in different ways and degrees, from serious disadvantage and harm" (Kahane and Savulescu, 2012, p.318). So their account claims that judgements about whether or not a person has been harmed should be based on comparisons between a person's state and what is statistically normal for a person like

[9] Shiffrin's account may also encounter problems in cases where the agent is irrational and in which the link between will and harm is not a reliable one.

her. On this account, for example, being blind is not statistically normal for humans (as opposed to moles or bats) and so a person is harmed if she is either blind or blinded. Furthermore, Kahane and Savulescu claim that their account of harm enables them to explain why a person is harmed if, as a result of a genetic condition, she is born blind. In this case, derived from Parfit's "non-identity problem" (1984), there is no foetus whose sight was potentially normal until being damaged. Moreover, any alternative foetus with the potential for sight would not be the same person, so this entails that there is no possible counterfactual comparative state of being sighted for the blind person. Purshouse (2015), however, objects to Kahane and Savulescu's use of statistical normality as the standard for judging whether or not a person has been harmed. He claims that Kahane and Savulescu would say a person is not harmed if they remain in a state that is above what is statistically normal for the species. So:

> "[w]hacking, say, a modern-day Michaelangelo over the head so that he could no longer paint something as great as the Sistine Chapel would not be to cause him harm under Kahane and Savulescu's theory provided he could still paint better than the average person" (Purshouse, 2015, p.7).

By introducing the idea of statistical normality to the concept of harm, Kahane and Savulescu have introduced a mechanism by which harms may be compared and ordered. The concept of statistical normality for particular states places them on a scale and so permits judgements that some states are worse than others. However, there is no scope under this system for comparing and ordering different types of state, say blindness and deafness, so Kahane and Savulescu do not say anything that may indicate how some harms may trump others unless these are the same types of harm but of a different severity. In other words, being partially sighted is less harmful than being blind, but there is no mechanism for comparing partial sightedness with partial deafness, in order to state which should take priority.

From my review of some representative accounts of harm, it has not been possible to infer anything about harms as trumps. The accounts aim primarily to explain harm itself and, since they do not set out to explain how harms may be prioritised, at best they only indicate some harms that may take priority over others. So, after this analysis, the question of whether some harms may trump others is left open. I now return to Feinberg, as his taxonomy of harms does suggest that some harms may be trumping ones.

5.6.6 Feinberg's Account of Conflicting Harms

As I have said, Feinberg defends a counterfactual account that relates harms to the thwarting of interests (which are desire-based[10]), such that an agent is in a worse position than they would otherwise have been. Furthermore, he categorises interests as being vital, extensive, or morally weighty ones (Feinberg, 1987, p.204), and each of these interests will have a corresponding harm. His reason for categorising interests is to help a legislator who wishes to balance different harms against each other. A legislator, Feinberg says, will have no problem deciding that killing and beating should be prohibited because of the harm to the victims. It is also straightforward, he continues, that walking and reading should be permitted because these are not harmful activities. However, there is a common type of situation that presents a problem for the legislator, and this is when

> "a certain kind of activity has a tendency to cause harm to people who are affected by it, but effective prohibition of that activity would tend to cause harm to those who have an interest in engaging in it, and not merely in the often trivial respect in which *all* restrictions of liberty (even the liberty to murder) are *pro tanto* harmful to the persons whose alternatives are narrowed, but rather because other substantial interests of these persons are totally thwarted" (Feinberg, 1987, p.203).

So, Feinberg says, for two people, A and B, who respectively have interests in X and Y, the legislator has to decide whether A's interest in X is more or less important than B's interest in Y. Feinberg admits that it is not possible to "prepare a detailed manual with the exact 'weights' of all human interests" (Feinberg, 1987, p.203). However he proposes relevant "dimensions" of interests and how these may be used to balance conflicting harms.

The first dimension of interests is vital ones. Thwarting of these interests "is likely to cause greater damage to the whole economy of personal […] interests than harm to the lesser interest will do, just as harm to one's heart or brain will do more damage to one's bodily health than an 'equal degree' of harm to less vital organs" (Feinberg, 1987, pp.204,205). So, Feinberg claims that a person's welfare interests—her interests in psychological and physical health—are the most vital, and if these interests conflict with another person's interests that are not vital then the former should take priority. An example of this is a recent case in which a person's vital interest in having accurate

---

[10] This is a contentious aspect of his account since an agent's interests may not necessarily track what is harmful to her.

information about the welfare of her missing child was balanced against newspaper readers' much less vital interest in the same child[11]. However, it is not difficult to imagine cases where similar vital interests are pitted against one another and a satisfactory resolution is more difficult to find, such as the example, earlier in this chapter, of dialysis and cancer treatment.

According to Feinberg, a further dimension of interests that should give them priority is the extent to which they are intertwined with other interests. Feinberg claims that "interests tend to pile up and reinforce one another" (Feinberg, 1987, p.205). He terms these interests "extensive" ones. Using the newspaper case above, a parent's interests in knowing about the welfare of her child extend into other interests, such as family, community and societal ones. In contrast, a newspaper reader's interest in knowing about missing people is much less extensive—it is almost exclusively restricted to the person herself—and should not be given priority. It is possible, however, that there may have been a spate of missing people in the reader's neighbourhood[12], and although they are still not as extensive as the parent's, the reader's interest in knowing about these may extend to an interest in her own safety. The earlier case, which balances an interest in avoiding death from renal failure against an interest in receiving treatment for cancer seems to involve extensive interests on both sides, though these may differ in nature. For example, the interests of people who are close to both patients may be affected, as may those of other patients with the same illness and those of the staff. So this harder case shows that the degree to which interests may extend is not very clear, and so may not always enable different extensive interests to be prioritised.

The third and last dimension of interests which Feinberg suggests should give them priority is the extent of their "inherent moral quality" (Feinberg, 1987, p.205). Feinberg's examples of moral qualities that he would not prioritise are an interest in knowing details of Brigitte Bardot's sex-life and the sadist's interests in causing other people to suffer pain. He describes these interests as being unworthy because of the values that they promote, and in these unproblematic cases it is hard to disagree with him. It may be harder in the sadist case, however, if the victim consents to being

---

[11] I am referring here to the tragic case of Milly Dowler, whose phone was hacked by staff at the News of the World after her murder. (UK Government, 2012)

[12] The Yorkshire Ripper case is an example of multiple missing persons in a defined area where people, especially women, may have a welfare interest in finding out from the media about how the situation was developing.

harmed, as in *volenti* type cases. But even in this situation Feinberg suggests that there is "a case against protecting their spawned interests, based upon their inherent unworthiness" (Feinberg, 1987, p.206). So, for him, the wider effects of permitting sadism weigh against its permissibility. Note here that Feinberg has to explain "inherent moral quality" because he links interests to desires, and thereby has to qualify bad desires such as being interested in Brigitte Bardot's sex life.

So, within each of Feinberg's three dimensions of interests—vital, extensive, and morally weighty—harms can potentially be prioritised against one another. What Feinberg does not do, however, is to weigh harms related to interests in any one of these three dimensions against harms in another of the three dimensions. So vital interests, for example, are not weighed against extensive or morally weighty ones, and so on. This shortcoming in Feinberg's account limits the progress that can be made in developing a taxonomy of harms in which some harms are trumping ones. However, progress can be made with the question of whether or not some harms may trump others by exploring parallel lines of reasoning, and below I defend a claim that arguments about certain types of need indicate a parallel line of reasoning that can be applied to harms.

5.7 Parallel Lines of Reasoning in Needs and Harms

There are three aspects of arguments about needs that suggest a parallel line of reasoning might be usefully applied to harms. First, after Megone (1992, p.12), people in different circumstances have a wide range of different needs, to which there are corresponding harms when they are not met. So, in Australia, fair skinned people need protection from the sun whilst, in the Arctic, people may need more sunlight, because of the sun's effects on vitamin D levels and sometimes on mood. The corresponding harms in Australia are sun-burn and skin cancer, whilst in the Arctic the harms are vitamin D deficiency and sometimes depression. A second aspect of arguments about harms that suggests a parallel line of reasoning is that, if resources are scarce, then needs can come into conflict with one another, just as harms can conflict if there are not the necessary resources to protect or rescue all the people affected. Third, resolution of conflicts between needs will depend on the priority to be assigned to these needs (which may depend on the degree of necessity) in the same way that resolutions of conflicts between harms to people will depend on the severity assigned to them.

My focus, now, will be on the third aspect of arguments about needs, the line of reasoning that enables resolution of conflicting needs by analysing their "necessity". Megone (1992) claims, first, that what enables needs to be prioritised is an account, following Wiggins, of different types of needs: needs are categorical if they are grounded in human nature, and instrumental if they are grounded in contingent desires of agents. Second, Megone's account provides an explanation of how needs-claims may have different forces, or normative power, thereby enabling them to be prioritised. The conclusion of the argument is that the claims of categorical needs should always take priority over the claims of instrumental needs. The insights gained from an analysis of needs allow us to argue that if categorical-type harms can be identified, which also have greater normative power, then they, too, should also always have priority over other types of harm, or be trumping harms.

Feinberg also defends an account of needs, and I defend a claim that his account differs in important ways from those of Wiggins and Megone. Feinberg's vital interests (1987, p.204), being welfare related, may correspond to categorical needs which are also grounded in human nature. So the most vital interests in Feinberg's account could have priority. However, there are at least two potential objections to this claim. First, as I have said, Feinberg's taxonomy only permits weighing of interests within each dimension. So some harms that are related to either extensive or morally weighty interests may have priority over vital type ones. Second, Feinberg's account is desire-based, but in Wiggins and Megone, as we will see, categorical needs are non-desire-based and have priority over desire-based instrumental needs. In reply to the first objection, Feinberg could claim, first, that extensive interests will not generally take priority over vital ones, and second, morally weighty interests collapse into vital ones. However, since Feinberg's account is desire-based it does not enable sharp divisions between the different harms, thereby enabling them to be prioritised. The needs based accounts of Wiggins and Megone are an advance on Feinberg since they do enable sharp divisions that permit prioritisation, due to categorical needs, as distinct from instrumental ones, not being based on desire.

5.7.1 Wiggins and Megone on Needs

As we have seen, Wiggins (1998) separates needs into two sub-types: categorical and instrumental. First, however, he identifies needs as what are necessary for a person, and quotes Aristotle's account of this in the Metaphysics:

> "We call Necessary (a) that without which as a joint cause, it is not possible to live, as for instance breathing and nourishment are necessary for an animal, because it is incapable of existing without them: and (b) anything without which it is not possible for good to exist or come to be, or for bad to be discarded or got rid of, as for instance drinking medicine is necessary so as not to be ill, and sailing to Aegina so as to get money" (Metaphysics IV, 4, 1015a 20ff quoted in Wiggins, 1998, p.25).

This quote from Aristotle suggests different types of things that are necessary. First, there are things that are necessary for the distinctive type of life that animal lives, and thereby these are needed so that it may live as the kind of animal it is, and second there are things that are necessary for some other purpose, such as getting money. So a human needs air and food in order to live *qua* human, and she needs to go to Aegina in order to get money. In the first case, air and water are fundamental to existence *qua* human—these are so-called categorical needs—but in the second case there is a further unspoken desire-based need—presumably the money from Aegina is needed, or elliptical, for some further purpose. More modern examples are needing a replacement washer for a tap, or some new rosin for a bow. These are examples of instrumental needs, since satisfying them is a means to some further purpose such as having a functioning tap (for washing and preparing food etc.) or being able to play one's cello (for pleasure and to entertain etc.). Aristotle's case of drinking medicine so as not to be ill could be indeterminate; it could be either categorical if the illness is a life-threatening one, or instrumental if it is a minor one.

Megone (1992) cites an objection, amongst others, by Barry (1965, pp.47-49), that both categorical and instrumental needs as described above are elliptical, since it is always possible to ask what the end is towards which the need is a means. So, for example, air, water and life-saving medicine are all necessary for life, but there could be a further end beyond the end of living itself. Perhaps the person who needs air, water and life-saving medicine is a musician for whom rosin is necessary for her to play her cello. However, it seems trivial to circularise ends in this way. It is better instead to break the circle of ends at the point where the end is life itself: living should come top in a hierarchy of needs since no other end is possible without life. A further way of prioritising needs is by assessing their practical or motivational force.

It can be seen from the examples above that the practical or motivational force of instrumental needs depend on the value to the person of the end towards which the need is directed. So an agent may place different values on washers, and rosin, and may prioritise rosin above the washer if she is due to perform in a concert. More importantly,

however, instrumental needs carry a different practical force when compared with categorical needs. Aristotle does not make categorical needs depend on whether the agent values them. In other words, there is no further goal beyond living itself, towards which categorical needs are directed. This means that a categorical need "does not therefore depend for its existence on the existence of such a goal, and thus if it has motivational force it is itself the source of that force" (Megone, 1992, p.18).

Before attempting to assign relative priorities to categorical and instrumental needs, I will look in more detail at categorical needs. Megone considers Anscombe's account of categorical needs for organisms, which claims that organisms of any type have needs of appropriate environments in order to survive: without them they cannot flourish. "The need of the organism [*qua* that kind of organism] directly and intrinsically provides a reason for it to have such an environment, not hypothetically on some further aim" (Megone, 1992, p.21). Megone then applies this account to humans which, being a type of organism, also have needs in order to survive. Megone lists these as "appropriate quantities of food, water, warmth, shelter, security and probably (given the kind of thing a human is) certain psychological requisites such as love. In the absence of these a human being is not able to be a good member of its kind, and thus […] not able to be what it is" (Megone, 1992, p.22). So if the value of an instrumental need is the value of the goal that the need is directed at, then the value that can be derived from categorical needs is the value of humans themselves. This last step enables us to prioritise categorical needs over instrumental ones. A greater weight can be accorded to categorical needs because they "reflect directly the value of each human being, whilst instrumental needs reflect the value of goals a human being may have" (Megone, 1992, p.26). Furthermore, though less importantly, categorical needs also have precedence over instrumental ones because a human has to meet her categorical needs in order to meet her instrumental ones.

5.7.2 Needs and Harms

From the line of reasoning I have given above for prioritising needs, it is possible to identify a parallel line of reasoning for prioritising harms. This line of reasoning about needs also reflects a different account of interests to Feinberg's, which is that only some interests—namely instrumental ones—are related to desires, and non-desire-based categorical needs should trump these. If this is the case, then we have reason to think that harms that are parallel to categorical needs may be trumping ones. As mentioned

above, there are some similarities between Feinberg's vital interests and categorical needs, which also supports the claim that harms related to vital interests may be trumping ones. Vital interests are interests in welfare, and a person's welfare is enhanced if she survives and flourishes. So a person's welfare interests may be realised if she has food, warmth, shelter, and the other categorical needs listed above. There are some vital interests, however, that are not categorical needs, since categorical needs are ones that are important for sustaining human life, rather than being more broadly welfare based.

There are further significant objections to using Feinberg's account of needs in support of an argument for trumping harms. Feinberg's taxonomy only weighs different vital, extensive or morally weighty interests against themselves. So it is possible that some harms related to either extensive or morally weighty interests may have priority over vital-type ones. There is nothing about extensive interests in and of themselves, however, that gives us reason to think that they should take precedence over vital ones. Feinberg's example of an extensive interest is a motorcyclist's interest in driving noisily through the suburbs. This interest may extend to the "interest of the cyclist's employer in having workers efficiently transported to his factory, and the economic interest of the community in general (including me) in the flourishing of the factory owner's business; the interest of the motorcycle manufacturers in their own profits […]" (Feinberg, 1987, p.205). It can be seen that the extensiveness of this interest, or any other, does not necessarily give it weight by virtue of its extensiveness. If, however, the extensive interests included some vital ones then these may give them priority; but it would be the vital interests that would be doing this prioritising work. Morally weighty interests, too, could take priority over vital ones. Feinberg gives morally "light" examples in his discussion rather than morally weighty ones—e.g. knowing details of Brigitte Bardot's sex life. If he had given morally weighty examples, such as beating or murdering Brigitte Bardot, it is likely that these would have been ones concerned with human welfare. So in the modified Brigitte Bardot example, morally weighty interests have collapsed into vital ones; and the status of certain harms related to vital interests as potential trumping harms has not been threatened by extensive or moral interests.

A second objection to Feinberg's account of needs is that it is mind-dependent, since it is based on the interests of the agent, and these are in turn based on the agent's desires. As we have seen, however, Wiggins and Megone's account of categorical needs is not mind-dependent. The implication of Feinberg's account of interests being mind-

dependent is that it is less likely to be supportive of trumping harms since a clear line cannot be drawn somewhere in a list of mind-dependent needs in order to prioritise some of them as trumping ones.

A potential objection to drawing a parallel line of reasoning about harms from the needs-related line of reasoning is that, whilst categorical needs have priority over instrumental needs, and categorical type harms should thereby have priority, we lack an instrumental type harm. In reply we can say that people may be harmed if their instrumental needs are thwarted. So if, for example, a person's interests in washers, cold cures or rosin are thwarted, then she will have been harmed, but these harms are less grave than if her categorical needs had been thwarted. So an instrumental interest is akin to an instrumental harm.

In sum, Wiggins and Megone's arguments about categorical and instrumental needs have provided a helpful insight into harms: there are some harms, such as ones connected with sustaining life, that may be trumping ones and which should thereby take priority over other harms. I now turn to another line of reasoning, from Dworkin, R. (1989), that may be useful for my purposes in identifying harms that may be trumping ones.

5.8 Dworkin's "Rights as Trumps"

A second parallel line of reasoning that may shed light on whether or not some harms may trump others is the defence by Dworkin, R. (1989) of a claim that rights are trumps. Dworkin's line of reasoning may be helpful to reasoning about harms because it deals in weighing of preferences, and preferences—which in one view are a mind-dependent, positive valuation of some state of affairs—can be considered as parallel to interests, i.e. something that we have a stake in and which, if thwarted, can result in harms.

Dworkin's defence of his claim about rights is based on a need to "insulate" people from oppressive, prevailing preferences in a utilitarian state. On his account, some agents may not realise their preferences if the preferences of other agents thwart them. In order to explain Dworkin's account of rights as trumps it would be helpful to restate Feinberg's formulation of the type of situation where harms are conflicting, and to apply Dworkin's reasoning to this case. Feinberg describes

> "a certain kind of activity [that] has a tendency to cause harm to people who are affected by it, but effective prohibition of that activity would tend to cause harm to those who have an interest in engaging in it […] because other substantial interests of these persons are totally thwarted" (Feinberg, 1987, p.203).

Dworkin's background assumption is that the relevant State in this case is operating under utilitarian rules. This entails a premise "that the community is better off if its members are on average happier or have more of their preferences satisfied" (Dworkin, R., 1989, p.153). So, by analogy with Feinberg, call the activity that has a tendency to cause harm X[13]; and assume banning X "would tend to cause harm to those who have an interest in engaging in it" (Feinberg, 1987, p.203). Dworkin's utilitarian State decides to ban X because this increases average happiness or allows more people to have their preferences satisfied. However, as a result of this decision, there are some people, call them A, who are harmed by having their interest in X thwarted. Although there are obvious exceptions, Dworkin would argue that A should be insulated against State action by having a trumping right to X.

Dworkin suggests two possible lines for his argument. First, he says that there may be an alternative ideal to utilitarianism which is important enough to permit X. However, he asserts that this argument would be pluralistic and unlikely to succeed. Instead he pursues a second line of argument which is that "further analysis of the grounds that we have for accepting utilitarianism as a background justification in the first place […] shows that utility must yield to some right of moral independence […]" (Dworkin, R., 1989, p.154). To show this, he imagines a corrupt version of utilitarianism that gives more weight to the preferences of some people than to the preferences of others. In this version, he describes a person called Sarah, whose preferences count for twice as much as everyone else's, and also a group of people—"Sarah lovers"—whose preference is that Sarah's (undoubled) preferences should count for twice as much as everyone else's. So, in this case, utilitarianism must claim both that no one is entitled to have more of her preferences fulfilled than anyone else, *and* that Sarah must have more of her preferences fulfilled than anyone else. This version, with its inherent contradiction, would constitute a corrupt version of utilitarianism since it would undermine its "egalitarian cast".

> "Utilitarianism must […] claim truth for itself, and therefore must claim the falsity of any theory that contradicts it […] But neutral utilitarianism claims […] that no one is, in

---

[13] Dworkin's example for X is a right to publish pornography. (Dworkin, 1989 p.154)

> principle, any more entitled to have any of his preferences fulfilled than anyone else is" (Dworkin, R., 1989, p.155).

If, Dworkin says, "utilitarianism in practice is not checked by something like the right of moral independence (and by other allied rights) it will disintegrate […] into exactly that version" (Dworkin, R., 1989, p.155).

Dworkin suggests that the contradiction can be resolved if utilitarianism is "qualified" by restricting preferences that exclude political preferences. The means of this restriction of preferences is *via* rights to political independence that are trumping ones: "[t]he right of political independence would have the effect of insulating […] those who are not Sarah from the preferences of those who adore her" (Dworkin, R., 1989, p.158). It is important to note here that Dworkin is talking about the right of *political* independence. The right to political independence protects people from the political preferences of people such as Sarah lovers. Dworkin's other, more formal, example of political preferences is of Nazis who have preferences about the rights of Jews to fulfil their preferences. Political preferences, such as those of Nazis or Sarah lovers, differ from moral preferences, such as those in Dworkin's example of a preference to ban the publication of pornography.

Dworkin claims that moral independence is relevantly similar to political independence. So, for example, people should have a trumping right "over an unrestricted utilitarian defence of prohibitory laws against pornography[.]" (Dworkin, R., 1989, p.158). He defends his claim by arguing that neutral utilitarianism should not have a view on whether or not some sexual practices are degrading. Neutral utilitarianism should also refrain from taking into consideration the preferences of people who feel that some sexual preferences are degrading.

If we accept Dworkin's argument and return, by analogy, to Feinberg's formulation, we see that some people, A, who would be harmed by having their interest in X thwarted as a result of banning by a utilitarian state, should be insulated against the State's ban by having a trumping right to X.

5.8.1 Objections to Dworkin

There are at least two limitations in Dworkin's argument seeking to defend political and moral preferences by giving them trumping rights in a utilitarian state. First, he has not been specific about the scope of the moral preferences. It can be seen from the example

that he gives—of pornography—that the type of moral preference he has in mind is not necessarily as serious as a preference that is parallel to a categorical need. However, we assume that he would support a trumping right that avoided interference with more significant preferences. Despite this assumption, Dworkin does not make it clear what would count as categorical preferences in his account. Second, Dworkin's argument is dependent on the contradiction that arises when the relevant state is an utilitarian one. So it is not clear if his argument would go through were it not an utilitarian one. Our interest here, however, is in harms that may be trumping ones when pitted against other harms.

5.9 Conclusion

In the case of an agent being harmed through being presented with an additional option and where failing to offer the option will result in harms to other agents, it is necessary to decide whether one set of harms should take priority over the other. This chapter set out to establish whether it is possible to weigh two sets of competing harms as a result of some harms trumping others. If one of the harms is a trumping one, then this would establish whether the additional option either should be presented or should not be presented. In this chapter, I have defended a claim that some harms are, indeed, trumping ones. That is to say, some harms, in virtue of their nature, should always take priority over other relevant harms. This is important to my thesis because I defend a claim in chapter six that the harm of wrongful death, should PAS be made permissible, should always take priority over the harm of unbearable suffering, should it not be made permissible.

The first step in defending my claim about trumping harms was to argue that the literature about both weighing harms, e.g. Taurek (1977), and harms themselves, e.g. Feinberg (1987), do not provide arguments that may allow us to infer that some harms are trumps. Following a parallel line of argument about needs, after Megone (1992) and Wiggins (1998), was more fruitful, showing that harms that are parallel to categorical needs are trumping ones and should, by their nature, always take priority over other harms. Ronald Dworkin's argument about "rights as trumps" (1989), although supportive of the idea that some harms may be trumping ones, did not tell us what counts as a trumping harm and is restricted to protecting rights in the face of a utilitarian regime. This argument, then, is clearly not applicable to the types of case that are my main focus.

The findings from the harms literature, which were unable to provide insights into how some harms may be trumping ones, are summarised below. First, the arguments about Taurek cases have not been decisively resolved and are, thereby, not useful for my purposes. In cases where the harms are lethal ones, Taurek (1977) argues that the numbers should not count. Furthermore, Taurek's argument implies that lethal harms may have a special status that should give them weight, even when they are balanced against a larger number of other lethal harms. As we have seen, however, his argument is based on difficulties aggregating claims to rescue on the two sides. It also reaches an unpalatable conclusion that demands a decision-procedure based on coin-tossing, even in numerically very imbalanced scenarios. After Lübbe (2008), a reply to Taurek from Kamm (2005), also fails to show how aggregation is possible. However, Lang (2005) offers a plausible strategy based on taking into consideration both selection and outcome fairness. In some scenarios coin-tossing may be fairest and in others the numbers should be counted, but there are also borderline cases in which it is not obvious which strategy to follow. So this still leaves unanswered the question of the best decision-procedure for Taurek cases.

There are also cases in which the harms on the two sides are of a different severity. Solutions for this type of case may rely, after Scanlon (1998), on agents on one or other side making a reasonable objection to the proposed principle on which selection is to be made. This procedure, too, cannot help in more balanced cases in which the agents on both sides may make a reasonable objection. Furthermore, this account is question-begging about what may be reasonable. Kamm's argument about irrelevant utilities (1993) offers little help with an argument about trumping harms since, in her account, harms that always give way to other ones are insignificant, and there is a need in my account to weigh different, non-trivial harms. Furthermore, Kamm's account does not aim to identify harms that should always take priority.

When more significant, but different, harms are balanced, Scanlon's "individualist restriction" (1998) is capable of offering guidance in some cases, but in others there may be a stalemate between significant (but different) claims to rescue, in which the agents on both sides feel they are entitled to help. So, this part of the literature on harms, too, does not give reason to infer that some harms may be trumping ones. I conclude, then, that it is not possible to infer from arguments about resolving conflict cases that there are some harms that are trumping harms.

Second, an analysis of selected writings on harms, themselves, did not shed light on how best to weigh them. However, these writers did not set out to prioritise harms, and merely indicated that harms may be different in severity and may also be weighed in certain restricted cases. Feinberg's taxonomy, which includes harms related to thwarting of vital interests, does, though, suggest that more vital interests should take priority over less vital ones. However, since Feinberg's account is interest (or desire) based, and thereby mind-dependent, it is less helpful in identifying harms that should always take priority.

In the remainder of the chapter, I turned to a parallel argument about needs, after Wiggins (1998) and Megone (1992), which helps defend my claim that certain types of harm trump others. Categorical needs should always take priority over instrumental ones, and as categorical needs are parallel to certain harms, so harms related to categorical needs are trumping ones. Dworkin's (1989) argument about rights as trumps was less helpful to me, however, as he is not clear either about the scope of the rights that he seeks to protect or how he may identify rights that are indicative of categorical-type harms.

I have argued in chapters one to four that an agent may be harmed if she is presented with an additional option, and that these harms may occur if she either has certain types of weak character or if the additional option causes a harmful context of choice. In this chapter, I have argued that categorical needs trump instrumental needs, and that the reasons for thinking this can be seen to apply in a similar way to the case of harms. I conclude, then, that certain types of harm trump other types of harm. In other words, harms related to categorical needs trump harms related to instrumental needs. Having established this, I am now in a position, in chapter six, to apply these arguments to the case of physician assisted suicide.

Chapter 6. Is the Permitting of Physician Assisted Suicide a Desirable Extension of Patient Choice?

6.1 Abstract

In the preceding five chapters, I have argued that an agent may potentially be harmed if she is offered an additional option and that, in addition to the types of harm identified by Dworkin and Velleman, there are two further types of harm, that of weak character and that of context of choice. I have also argued that these harms may be weighed against the harms to other agents if the same option is not offered, and that some harms trump others. This chapter applies these arguments to the case where Physician Assisted Suicide is made an additional option for selected patients. I defend a claim that patients, first, with certain types of weak character, and second, as a result of the context of choice, may be harmed as a result of wrongful death should they be offered PAS. Furthermore, this harm is a trumping one and should, thereby, always take priority over non-trumping harms such as the harm of unbearable suffering. Importantly, despite its significance and the limitations it imposes, unbearable suffering is not on a par with wrongful death since it does not, first, prevent patients from functioning autonomously in the context of requests for PAS, or second, prevent patients from meeting their psychological needs for survival. I conclude from this that the permitting of PAS is not a desirable extension of patient choice.

6.2 Introduction

I argued in chapters one to four that an agent may be harmed if her choice is extended through being offered an additional option, if she has certain types of weak character and/or the resulting context of choice has a particular configuration. In chapter five, I drew a comparison with needs (Wiggins 1998, and Megone 1992), in order to argue that some harms are trumping ones and should always be prioritised against non-trumping harms.

I now defend a claim in this chapter that the permitting of PAS is not a desirable extension of patient choice. In order to do this I will apply the key arguments in chapters one to five to cases where PAS is made a live option. I will argue that, should PAS be made permissible, patients to whom it is offered may suffer wrongful death. I

will defend, too, how the harm of wrongful death trumps the harm of unbearable suffering of patients for whom the offer of PAS is not made.

Applying the arguments about character to cases of patient choice, should PAS be made permissible, I claim, first, the types of weak character that I have defended, namely *acrasia* (Aristotle, 2002), undue self-deprecation (Richards, 1988) and undue lack of confidence in judgements, may influence a patient to choose PAS and, thereby, suffer wrongful death. Her death would be wrongful because she may choose PAS when this is not what she would have chosen before the offer of PAS was made, or because her choice reflects a vice of character. In addition, I argue that, in virtue of the way character is formed (Aristotle, 2002), a patient who is dying may have a character that is inchoate, or incompletely formed, in respect of making some choices, such as accepting an offer of PAS. I claim that an inchoate character may increase the chance of a dying patient behaving *acratically*, but I am agnostic about its effects on undue self-deprecation and undue lack of confidence in judgements.

My second claim is that a patient in a situation where PAS is made permissible may be harmed by the context of choice. Should a patient choose PAS as a result of being influenced by features of the context of choice that should not have been salient to her, and/or not being influenced by features that should have been salient to her, in order to make a rational decision, then she may consequently be harmed by suffering wrongful death.

Should PAS not be made permissible, however, then certain patients who are suffering unbearably at the end of their lives, and whose suffering is unrelievable, may be seriously harmed because they do not have the option of ending their suffering by dying. I argue that in order to decide whether PAS should be made permissible, the harms that will accrue to these patients if it is not made permissible have to be weighed against the harms that will accrue to certain different patients if it is made permissible. I claim that the harm to patients who die wrongfully as a result of choosing PAS is a trumping harm over the harm to patients who are suffering unbearably at the end of life and who are denied PAS. One important objection here is that, in virtue of its nature, unbearable suffering is on a par with wrongful death since it may prevent patients from acting autonomously: so unbearable suffering cannot be trumped by wrongful death. However, despite being severely constrained, patients who are suffering unbearably are able to act autonomously; this is how they may communicate their suffering and consequent

choices, including a potential choice of PAS or other options that may relieve suffering. Furthermore, patients who are suffering unbearably are able to meet the psychological requisites for human survival in virtue of having personal relationships (Blum, 2008). These relationships normally entail a compassionate response towards the patient with the potential result of benefits to her. Since the harm to people who die wrongfully should, thereby, always take priority, I conclude that the permitting of PAS is not a desirable extension of patient choice.

Before applying my arguments to the case of PAS, it will be helpful to give some background to PAS and to the recent Act presented to the House of Commons (2016). since this sheds light both on the types of situations where PAS may be offered and on some of the pre-conditions for offering PAS. As we will see, both the types of situation and the pre-conditions for offering PAS are relevant to my arguments.

6.3 Background to PAS and the Act Recently Presented to the House of Commons

Should PAS be made permissible then this would extend choice for some patients. So, if the Assisted Dying (no. 2) Bill (2016)[1] recently presented to the House of Commons by Mariss were enacted, a patient could be offered PAS, but only if she met the following criteria: she must be aged over eighteen, have a terminal illness with a prognosis of dying within six months, and be a resident of England or Wales. Importantly, too, in order to protect vulnerable people from wrongful death, a patient must also have "a voluntary, clear, settled and informed wish to end his or her own life" (2016, p.1) before they may be provided with assistance to do so. In order to defend my claim that patients who meet the requirements of the act may, nevertheless, wrongfully die[2], I must argue, first, that an apparently "clear, settled and informed wish" may be

---

[1] There has been a move towards using the term "assisted dying" instead of PAS. This may have been an intentional move on the part of campaigners for PAS to clarify that assisted dying is intended to help patients who are near to death rather than people who do not have a terminal illness. The change in term also removes the word "suicide". Suicide is seen as a pejorative term since it is has negative associations with mental illness. I have used the term PAS in this dissertation in order to be consistent with earlier literature, but each time I write PAS this could be replaced with "Physician Assisted Dying".

[2] Supporters of PAS claim that it is possible to build adequate safeguarding measures into any provision of PAS, in order to protect vulnerable people from requesting PAS. So, according to them, it would not be possible for anyone to wrongfully die under the conditions stipulated in an Act of Parliament that legalises PAS. My arguments based on the harms of weak character and context of choice refute this claim, which is discussed in more detail in the Appendix.

the wish of a patient who, before the offer of PAS is made, did not in fact wish to die prematurely or be a wish that reflects a vice of character. I will use my account of types of weak character to defend this part of my argument. Second, even when voluntary, a patient's wish to end her own life may reflect aspects of the context in which she is choosing. So I will also argue that changes to the context of choice, through the addition of the option of PAS, may influence a patient to select PAS when she would not have done so previously, thereby resulting in her wrongful death.

6.4 Weak Character in the Case of PAS

In this section I defend the claim that three types of weak character—namely *acrasia*, undue self-deprecation and undue lack of confidence in judgements—may result in a patient selecting PAS and suffering wrongful death. *Acrasia* may result in a patient selecting PAS when this is not what she would have selected for herself before the offer of PAS was made, and undue self-deprecation and undue lack of confidence in judgements, as vices of character, may influence the patient to choose this option when it is not the best one for her. Furthermore, inchoateness of character in end of life situations, may increase the chance of a patient acting *acratically*. This is because the patient, first, may lack appropriate knowledge that is relevant to her end-of-life situation, second, she may be subject to extremes of emotion and, third, she may be choosing in a situation where a quick decision is necessary. I am agnostic, however, about the effects of inchoateness of character in end-of-life situations on a patient's undue self-deprecation and undue lack of confidence in judgements.

6.4.1 *Acrasia* and the Option of PAS

We have seen in chapters two and three that an agent who is offered an additional option, along with a pre-existing option, normally deliberates about these in order to make a choice between them. The agent will normally also experience an emotional response to both options (Aristotle, 2002). In a simple case, the *acrates*—call her Julie—acts in line with her irrational emotions/desires[3] and against her rational choice by selecting the option towards which her irrational emotions have taken her. Julie is harmed by her selection because she has chosen the option that she reasons will be best for her, but then acts against this choice and moves in line with her irrational emotions

[3] As in chapter three, I will use "irrational emotions" or just "emotions" for irrational emotions/desires.

towards the alternative option. If the additional option being presented to Julie is PAS, she may reason that she would rather not choose PAS and remain alive. Nevertheless, Julie may select PAS if her irrational emotions have taken her towards this option. In this case, Julie has wrongfully died since, she has chosen to end her life despite reasoning that staying alive would be best for her. Note here, that Julie already has an *acratic* tendency, and presenting an additional option of PAS has enabled this tendency to be exercised.

We can look in more detail at Julie's emotions and feelings by drawing from the account of *acrasia* in chapter three. We saw there that pleasure, pain and anger are key feelings with relevance to *acrasia* (Aristotle, 2002). So Julie may have been drawn towards PAS because she has an irrational desire for the relative "pleasure" of being dead, in contrast to the suffering she experiences in life. Despite this being a difficult argument to sustain because death cannot be known (Nagel, 2013), Julie may have a *conception* of dying or being dead as pleasurable, perhaps as a peaceful sleep, and it is this conception that may have taken her towards PAS. Death may also seem "pleasurable" in contrast to the pain, physical or psychological, associated with her life, which may also push her towards PAS. Irrational anger, or irrational fear, perhaps of living with unbearable suffering, may, too, defeat Julie's reasoned choice to stay alive and push her towards PAS. Impetuosity (NE 1150b27-29), with its associated lack of deliberation, may also be relevant to the greater role of irrational emotion in Julie's decision. Impetuosity is perhaps more likely in end of life situations where the life expectancy is short. In this case, Julie may feel that she must act quickly in ending her life, before she deteriorates and becomes unable to exercise her options. Two further, important aspects of this account of *acrasia* in the case of PAS are, first, that the options in front of patients in end-of-life situations may plausibly arouse strong emotions and, second, that the novelty of the option of PAS for a patient may increase the role of irrational emotions in her decision[4].

I have argued in chapter three that an additional option that arouses strong irrational emotions may be more likely to influence how an agent acts than one that arouses weak irrational emotions. Consider a patient—call her Claire—who may be prone to oscillation between *acrasia* and *enkrateia* (where she acts in line with her choice and

[4] I analyse the effects of being offered a novel option in the case of *acrasia* in a later section, along with an analysis of the effects of being offered a novel option on my other two types of weak character.

against her irrational emotion). As I have said, an agent's character is not normally completely stable, thereby making possible this oscillation between character states (Annas, 1993, p.50). Claire may be swayed towards acting *acratically* if she finds herself in a situation, such as being in the grip of a terminal illness, that generates strong irrational emotions. So if she has reasoned that she would rather remain alive, but acts in line with her strong, irrational emotions and selects PAS, then this may be because the strength of her irrational emotions swayed her towards acting *acratically*. Alternatively, Claire may have acted *enkratically*, i.e. in line with her choice and against her emotions, if her emotions had been weaker. So, in the first version of this example, being faced with an option that arouses strong emotions, namely PAS, has resulted in Claire selecting this option contrary to her reasoned choice, and in so doing she is harmed through wrongful death.

6.4.2 Undue Self-deprecation and the Option of PAS

I now turn to the second type of weak character that I claim may influence a patient to select PAS—undue self-deprecation. As I have said, undue self-deprecation is one of two vices, with arrogance, in the area of an agent's assessment of her self-worth, abilities and entitlement. Since undue self-deprecation is a vice, choices that reflect an unduly self-deprecating character are ones that are bad for the agent. The virtue in this area, after Richards (1988), is humility; and humility entails an agent having an accurate conception of her self-worth. So, if an agent has an unduly self-deprecating character then she will have a conception of her self-worth that is lower than one that would be justified in a fair assessment. Importantly for my purposes in defending a claim that offering PAS may be harmful, an agent may be more prone to exercising the vice of assessing her self-worth as being lower than it should be in a fair assessment, and consequently making a harmful choice, when she is presented with the additional option of PAS.

Consider Maisie, who is terminally ill and who has become increasingly dependent on others for her care. Maisie is a widow whose adult children have been helping her with various activities of daily living, such as shopping, cooking and housework. However her children live at least twenty miles away and each has children of their own. Maisie is unduly self-deprecating, and so has a lower than justified self-assessment of her worth and entitlement. She does not believe that her continuing existence has worth—she does not feel she is entitled to ever-increasing assistance with her daily activities

and believes she is a burden on her family. Maisie's self-assessment is lower than justified because her children value her company, feel that she is entitled to their visits and care and derive pleasure from caring for her and supporting her at home: they are prepared to do whatever is necessary to care for her until she dies.

Maisie's condition deteriorates, and her GP and Consultant now feel that she has less than six months to live. Since she is now eligible for it, she is offered the additional option of PAS. Maisie must now make a decision about whether she should either opt for PAS or for continuing palliative care and support from her family. In this situation, Maisie's sense of her own worth and entitlement is relevant to her decision whether or not to choose PAS. If she does not feel that she is entitled to care and support, and if she feels that she is a burden on her family, then she may feel that PAS is a suitable option for her. Maisie reasons that if she chooses PAS and dies, then her family will no longer have to care for her—they will be spared up to six months of burdensome caring. In this case, Maisie has been harmed by the offer of PAS because this enabled the manifestation of her vice of undue self-deprecation, resulting in a harmful choice to die prematurely, and thereby wrongfully.

Should Masie not have had the option of PAS, then she may have continued to live until she died from her illness, sometime within the following six months. She may have had to endure distressing feelings as a result of perceiving that she was not entitled to care and was a burden on her family, but would have had the chance of these being mitigated by their reassurance and loving care. Furthermore, since character traits are not completely stable, aspects of Maisie's character might have changed so that she could become more able to virtuously appreciate that she had worth and was entitled to care and support.

6.4.3 Undue Lack of Confidence in Judgements and the Option of PAS

The third and final type of weak character that I claim may influence a patient to select PAS and suffer wrongful death is undue lack of confidence in judgements. I defended a claim in chapters one and three that undue lack of confidence in judgements may result in harm to an agent who is offered an additional option. I also argued that this harm is distinct from Dworkin's potential harm from an additional option that he terms "responsibility for choice" (Dworkin, G., 1988, p.67), although it may *additionally* occur in cases used by Dworkin to illuminate his harm. In chapter three, I defended an account of how this additional harm may occur: an agent may suffer the harm of

responsibility for her choice in the context of a contrary tide of opinion to her own. Furthermore, she may let go of her values as a result of an undue lack of confidence in her judgements about them. I also argued that, if an agent is overconfident in her judgements about her values, then she may be unwilling to let go of them, even when they are clearly misplaced. The agent may then choose in line with these values and later regret her choice[5]. This latter confidence-related harm, however, is clearly not applicable in the case of PAS where a patient dies wrongfully. So, it is the vice of undue lack of confidence in judgements that may adversely affect the choice made by an agent who is offered an additional option. I now turn to how undue lack of confidence in judgements may influence a patient who is offered the additional option of PAS.

Consider Nadeen who, in the now familiar story, is presented with the additional option of PAS. Nadeen has Motor Neurone Disease (MND) and her values indicate that she should not choose PAS. She is well-supported, in receipt of palliative care and would like to live as long as possible, despite her illness. However, Nadeen has undue lack of confidence in the judgements that flow from her values as they relate to PAS for people with Motor Neurone Disease. She thinks her judgement, arising from her values, that suggest to her that she should not request PAS, may not be well-founded and she wonders if, instead, she should opt for PAS. Furthermore, Nadeen lacks confidence in her judgements arising from her values, despite the fact that she has thought deeply both about dying from Motor Neurone Disease and the implications of accepting PAS. She should thereby feel more confident about her judgements than she does. So, if Nadeen does opt for PAS, then, all other things being equal, she will suffer wrongful death. If she had had an appropriate degree of confidence in her judgements flowing from her values, then she may not have selected PAS and would have continued to live.

It can be seen in Nadeen's case that her undue lack of confidence could combine with other features of the choice situation that are part of her context, such as highly publicised cases that are relevant to her. However, Nadeen's vice of undue lack of confidence in her judgements does not necessarily require pressure in order to be manifested. Before I turn to the context of choice and its harmful role in influencing patients to choose the option of PAS and thereby suffer wrongful death, I argue that

---

[5] Harman (2009) has argued that if an agent feels she will be glad if she acts a certain way, then this gives her a good reason to think that she should act that way. In my case of overconfidence in which the agent later regrets her actions, she does not have sufficient insight into her overall situation to reason in this way: she feels she will be glad if she acts in line with her overconfident character.

there is a further aspect of character that is relevant to end of life situations. In virtue of the way character is developed, the patient's character at the end of life may be inchoate with respect to choices related to dying. This feature of character may increase the chance of a patient behaving *acratically*, but has indeterminate effects on undue self-deprecation and undue lack of confidence in judgements, when a patient faces end-of-life decisions.

6.4.4 The Relevance of Formation of Character in the Context of Dying

I argue here that the Aristotelian account of character that I defended in chapter two has general implications for end-of-life situations. In brief, my account suggests that character may not be completely formed in respect of end-of-life decisions, as the patient's knowledge of her novel situation may be limited. If character is incompletely formed then this may, first, increase the chance of the patient being *acratic*, and of her reasoning being weakened, in respect of her end-of-life choices. However, second, there is no reason to assume that the vices of undue self-deprecation and undue lack of confidence in judgements are more likely to be manifested in patients facing novel situations such as end-of-life ones.

The account of character formation that I defended in chapter two incorporates fundamental roles for emotion, reason and habituation. In this account, character is a disposition of an agent's soul such that she experiences emotions and reasons and chooses in certain ways that are shaped by habituation during the course of her life (Aristotle, 2002). An agents is born with pre-dispositions—tendencies to experience certain psychological states (Hursthouse, 1999)—but how these are shaped into character will depend on the environment the agent finds herself in and how she experiences this.

Some situations in an agent's life, such as meeting someone new, are commonplace, and over the course of her life an agent will experience desires and emotions and make rational or irrational choices in relation to this type of experience. Some of the agent's desires and emotions, and corresponding choices, are more likely to recur in similar situations and some are less likely to recur, thereby shaping her character through habituation. The agent's character may also be habituated under guidance from another person (Aristotle, 2002). So, if an agent repeatedly encounters similar situations, her previous choices and actions and her assessments of these may shape the way she chooses and acts in similar situations in the future. In broad terms, the agent may take

pleasure or pain in her choices and actions. If she takes pain in them then she may choose and act differently in a similar situation in future, and if she takes pleasure in them, then she may choose and act in a similar way in future. The agent will thereby tend to experience similar emotions and make similar choices in situations that she has repeatedly encountered and which are familiar to her. Despite becoming increasingly stable, however, her actions and choices will not be completely predictable (Annas, 1993, p.50).

The corollary of this account of the manifestations of an agent's character in commonplace situations is that in situations that are novel, the agent's emotional responses and choices may be inchoate, or less stable. It is in the nature of dying that it is a novel situation: an agent will not experience it on more than one occasion. The agent's desires, emotions and choices will not, then, have been guided and habituated, and her character will not be formed with regard to dying. It is likely, however, that the agent will have experienced the deaths of others and may have witnessed their emotions and choices. She may also have learnt about how people have experienced dying, for example through literature or film. There are public ventures[6] that aim to familiarise people with death and dying so that they may be better prepared for their own deaths. Learning about dying from these sources may influence an agent's beliefs about death, but this is not the same for her as having previous experience of relevant desires, emotions and choices when she, herself, is faced with dying.

One case, however, in which a patient may experience the prospect of dying on more than one occasion[7] is if she has a terminal illness, such as cancer, where the illness may go into remission after treatment[8]. The patient may believe that she is dying and embark on potentially life-saving treatment, but in the knowledge that it may not be successful. Remission may be followed by relapse and the patient may decide to undergo a further course of treatment. This cycle may repeat itself several times, and on each of these occasions, since there is a chance that the relapse will be a fatal one, the patient is

---

[6] The National Council for Palliative Care set up "Dying Matters" in 2009, with the aim of helping "people to talk more openly about dying, death and bereavement, and to make plans for the end of life."

[7] Further examples of cases where people are able to experience dying more than once may be provided by soldiers, and perhaps also by people who engage in dangerous sports such as wingsuit-flying or free-climbing. That is not to say that both these groups of people necessarily believe they are putting themselves in lethal situations: they may deny that they are doing so.

[8] I am grateful to Ilana Gluck for alerting me to this type of case and its consequences for character formation.

experiencing repeated episodes of dying. So there is the possibility in this type of situation for a patient to habituate her emotions and choices, and in doing so to shape her character in relation to her end-of-life situation.

One objection here is that these types of episode occur in a relatively compressed period at the end of life and are not experienced throughout a lifetime such that they have a more significant effect on an agent's character. Furthermore, terminal illnesses, such as cancer, become more prevalent with increasing age, so patients who experience repeated relapses are on the whole older. As a result of their advanced age, these patient may be less susceptible to the habituation of new dispositions. On the other hand, there are some conditions, such as cystic fibrosis, in which premature death is a possibility that hangs over the patient's entire life. For many years before her death, the patient's life threatening illness may require life-saving treatments, some of which, such as heart and lung transplantation, may themselves be life-threatening. So a patient who has cystic fibrosis may undergo some habituation of her desires, emotions and choices in respect of dying. Notwithstanding these examples, most people do not have the opportunity to undergo habituation in this way, and this may have an effect on their end-of-life character. I consider, now, the implications of this account of character at the end-of-life for my three types of weak character as they apply to an offer of PAS.

6.4.4.1 *Acrasia* in the Context of Dying

There are aspects of *acrasia* that make it *more* likely in novel, end of life situations. In a novel situation a patient's ability to choose may be impaired because she cannot "grasp information" about the particulars of the situation with which to deliberate and choose (Manson and O'Neill, 2008, p.5). This type of case contrasts with alternative ones in which the options are familiar to the agent and in which she may have previously deliberated and chosen what to do, so that she has a good knowledge of the relevant particulars. In a novel situation, then, irrational emotions may be more likely to influence an agent's action because her choice may be undermined either, first, by lack of knowledge, or second, because either she lacks knowledge or her knowledge is less likely to be actualised so as to make "a difference to the world" (Broadie, 2002, p.386).

Consider Claire again, but in a situation where she is dying from cancer and suffering from unbearable pain. Her options, following the additional choice of PAS, are palliative treatment, a course of experimental chemotherapy with uncertain side effects, and PAS. In a first variant of this case, Claire has not previously been offered the

experimental chemotherapy or palliative treatment, so her knowledge about them and their comparative benefits and risks is incomplete. She is frightened about potential side effects of the treatment and also the possibility that palliative care may not relieve her symptoms, so she selects PAS. In this first variant, lack of knowledge about the chemotherapy and palliative care has allowed her fear about side effects and unrelievable symptoms to push her towards the option of PAS, despite her previously settled desire not to choose it. In a second variant of the case, Claire reasons that the chemotherapy would be best for her, despite not knowing what it entails, and chooses this instead of PAS. Because her knowledge is incomplete, however, her choice fails to be actualised (Broadie, 2002, p.386). In these two variants, it can be seen that a patient may be more likely to behave *acratically* when making end-of-life decisions, due to having impaired knowledge about the options.

6.4.4.2 Undue Self-deprecation and Undue Lack of Confidence in the Context of Dying

I will now argue that the vices of undue self-deprecation and undue lack of confidence in one's judgements are both pervasive character traits that may manifest themselves in a range of relevant situations including end-of-life ones. First, self-deprecation, as I have said, is a vice in respect of self-worth, abilities and entitlement. It seems unlikely at first blush that an agent who is unduly self-deprecating in one sphere will be more virtuous in respect of self-worth in another one. Consider James, who has lost his job following an injury in the work-place and is now on benefits. He is unduly self-deprecating and does not contest a decision to reduce his benefits, despite his friends telling him that the decision seems unfair. Throughout his life, James has been unduly self-deprecating: despite being a very good player, he did not feel entitled to a place in the school football team; he does not pursue people who have short-changed him. Whatever the underlying reasons for his undue self-deprecation, it seems likely that he will also be unduly self-deprecating in respect of his entitlements in an end-of-life context. This is because undue self-deprecation is applicable across a range of different situations that are relevant to an agent's abilities and entitlement.

On the other hand, consider Janet, who has been much less self-deprecating than James over the course of her life and has always fought for what she felt she was entitled to. It is likely that Janet will not be unduly self-deprecating in an end-of-life situation as, although the situation she faces is new to her, she has always pursued her entitlements

and the stakes are higher than at any other time in her life. However, it is possible that Janet may be more unduly self-deprecating in an end-of-life situation since, as I have said, an agent's character is not completely stable. This possibility seems less likely than the possibility that James will become less unduly self-deprecating and stick up for his rights to proper care whilst he is dying, or refuse the option of PAS if he is offered it.

I now turn to the vice of undue lack of confidence in one's judgements in an end of life situation. Like undue self-deprecation, undue lack of confidence in one's judgements is applicable across a range of relevant different situations. The situations in this case are ones in which the agent holds values but unduly lacks confidence in the judgements that flow from them. Consider Adam, who has the vice of undue lack of confidence in judgements about his values. Earlier in his adult life, Adam found it especially difficult to take responsibility for choices he made. These were choices that were in line with his values, e.g. to sell his car in order to reduce his carbon footprint, but he was unduly lacking in confidence about the judgement that flowed from his values and this lack of confidence was harmful to him. In a variant of this case, Adam has the same vice, but lets go of his values as a result of undue lack of confidence in his judgements arising from them—in this case, he is harmed by not selling his car in order to reduce his carbon footprint.

Now consider both versions of Adam in an end-of-life situation. They both have the vice of undue lack of confidence in their judgements flowing from their values, and when they are presented with the additional option of PAS, this presents them with the opportunity to exercise their vice. They act in the same way as Nadeen did in section 6.4.3.

Since character is not completely stable it can be called "indeterminate" (Vranas, 2005). However, in the light of the account of character that I have defended it is more likely than not that a patient will behave in line with her character in an end of life situation. The conclusion I reach here is that since character is not completely stable it is not possible to predict how a patient will behave in an end-of-life situation. However, not being able to make predictions about character is a different matter from giving up on the idea of character itself and the implication inherent within it that a patient is more likely than not to behave in line with her character when faced with end-of-life decisions.

I have previously argued that if an agent is aware of the views of other people in a choice situation then this may also influence her choice. The views of other people in a choice situation are part of the context of choice, and I now turn to the context of choice and its relevance to an offer of PAS.

6.5 The Context of Choice in the Case of PAS

I argued in chapters one and four that the context of choice is a further potential harm that an agent may experience when she is offered an additional option. As a result of the context of choice, first, she may perceive as salient a feature of the choice situation that she should not find salient, e.g. the relative valuation of the cross pen and the cheaper pen in the "decoy" case (Tversky and Simonson, 1993). Second, she may not perceive as salient a feature of the choice situation that she should perceive as salient, e.g. if the feature has been presented in a particular way that has the effect of lessening its salience, such as the rash in the poorly lit examination room from the medical analogy. Each of these two instances of changes in salience attributable to the context of choice may alter the choice an agent makes to one that she would not have made before the additional option was offered. Before the additional option was offered, she would have selected an option based on what would be rational for her to choose. If the agent's choice is not one that she would have made before the additional option was offered, then this may be harmful to her. I now defend a claim that in the case of a patient who is offered the additional option of PAS, the resultant change in the context of choice may influence her to select PAS when this is not what she would have selected before the offer was made. So she will, thereby, be harmed as a result of wrongful death.

In chapter four I drew on three types of case from the social sciences literature that shed light on the way in which the context of choice may exert a harmful influence. These are framing (Tversky and Kahneman, 1986), decoys (Tversky and Simonson, 1993) and positioning (Thaler and Sunstein, 2008). In these types of case the offering of an additional option (decoy cases) or the way in which the option is presented (framing and nudge cases) may have an influence on the salience to an agent of features of the choice situation, and thereby how she makes her choice. Changes in the salience to an agent of features of the choice situation result in her either finding salient features of the choice situation that she should not find salient, or not finding salient features of the choice situation that she should find salient. I now apply these three types of case in a choice situation where a patient is being offered PAS.

Framing is the first of the three types of case that illustrate the harmful effects of the context of choice. Consider a patient who is offered the additional option of PAS alongside the option of palliative care. These two options can be presented to a patient in two different "frames". First there is a survival frame and second there is a suffering frame. In the survival frame, PAS may appear worse to a patient since only the palliative care option results in survival. In the suffering frame, palliative care may appear worse to a patient than PAS since there is a chance that palliative care may not be effective in relieving suffering, whereas PAS ends suffering as a result of death. Now consider a patient in this situation who has a preference not to select PAS but who is presented with the option of PAS within the suffering frame. PAS may be regarded as a means of relieving suffering, so it is likely to be framed in this way. The patient now finds the apparent beneficial effects of PAS, when it is offered within this frame, salient to her. Before PAS was offered to her she would not have found it salient in helping her to make a rational choice. So, as a result of being offered the option of PAS within the suffering frame, and selecting it when she would not previously have selected it, the patient is, then, harmed as a result of wrongful death.

In the second of the three types of case from the social sciences literature, the decoy effect, an agent is initially offered two options, e.g. "$6 and an elegant Cross pen", and makes a choice between them. An additional option, e.g. "a second less attractive pen" (Tversky and Simonson, 1993), is then offered alongside the pre-existing ones and the agent is again asked to make a choice. The decoy effect demonstrates how the agent's initial preference, e.g. for the money when initially offered two options, may change when an additional option is offered. This change occurs because the agent finds the difference in valuation of the two pens more salient to her than her preference for the money.

The decoy effect can also be applied in the case of an offer of PAS. Consider a patient with advanced malignancy who has two options—chemotherapy or palliative care. Neither of these two options differ in respect of the suffering they may entail, but the chemotherapy offers a chance of extending the patient's life. The agent in this situation prefers the option of chemotherapy because she wants to live for longer. Now consider the same patient, but this time in a situation where she is offered PAS in addition to the previous two options. The patient now makes a favourable comparison between PAS and palliative care in terms of their effects on suffering—PAS offers immediate relief from suffering. Furthermore, the patient does not make a comparison between PAS and

chemotherapy since in this case chemotherapy does not entail any degree of suffering. In the light of the favourable comparison between PAS and palliative care, the patient now selects PAS rather than chemotherapy. In the second choice situation, the salience to the patient of the favourable comparison between PAS and palliative care has reduced the salience to her of the effects of the three options on survival. In the former, two-way choice situation, survival was the factor that lead her to select chemotherapy. In this example, then, the patient chooses PAS and dies, despite being more concerned about survival before the offer of PAS was made. She has, thereby, suffered the harm of wrongful death.

It could be objected that the way I have set up this PAS case differs from the money and Cross pen case. The decoy in the latter case is the inferior pen, and in the PAS case the additional option is not inferior in terms of suffering. However, what I claim to have established here is that despite survival being more important than suffering to the patient in my example, she chose PAS as a result of the favourable suffering comparison with palliative care. Perhaps the difference in gains from chemotherapy and palliative care in terms of survival were marginal, and this contributed to the effect of the offer of PAS on her choice.

The last of the three types of case that illuminate the harmful effects of the context of choice is “nudge” (Thaler and Sunstein, 2008). In nudge cases, the way in which the options are presented to an agent affects the salience to her of those options. Furthermore, the change in salience to an agent of the options may result in her selecting an option that she would not previously have chosen before the presentation of the options was altered, e.g. healthy food in the cafeteria case described by Thaler and Sunstein. In the case of PAS, then, a patient who would not previously have chosen PAS may select it and die wrongfully as a result of the way in which the option of PAS is offered.

Consider a patient, Nadia, who is suffering in the course of a terminal illness and who is offered PAS. She has had some palliative care but this has not alleviated all of her symptoms and she visits her doctor to discuss her options. These options include PAS but, importantly, Nadia has previously expressed a view that she would not be willing to consider PAS in her situation.

Clearly there is no direct, spatial, "cafeteria" equivalent in the way that PAS may be offered to Nadia, but there are associated features of the way the offer is made that could alter its salience to her. A nudge is "any aspect of the choice architecture that alters people's behaviour in a predictable way without forbidding any options or significantly changing their economic incentives" (Thaler and Sunstein, 2008, p.6). So, in this case, Nadia's behaviour may be changed if the doctor discussing all the options available to her were to act according to his impression that she is suffering unbearably by giving more prominence to the option of PAS than the other options. Perhaps he also feels that the remaining palliative options may not be completely successful. The doctor in this case would be acting in a similar way to Carolyn, the hypothetical director of school food services in Thaler and Sunstein's cafeteria, who makes a judgement about what she feels is best for her pupils.

The doctor in Nadia's case could be criticized for acting paternalistically. This, however, is a disputed aspect of nudge (e.g. Hausman and Welch, 2010) which I do not need to resolve. It is necessary for my purposes only that I establish that a patient may be influenced by changes, such as those entailed in nudge, in the context of choice. If Nadia's doctor gives undue prominence to PAS and she chooses it despite previously expressing a view that she would not choose PAS, then Nadia may be harmed through wrongful death.

A further way in which Nadia may be harmfully influenced by the context of choice is if she is aware of a tide of opinion, contrary to her own, that people in her situation are better off if they choose PAS. If Nadia has Motor Neurone Disease, she may be aware of recent cases of people with Motor Neurone Disease who have died as a result of PAS and whose cases have been publicised[9]. Her awareness of these cases may alter the salience to her of PAS, with the result that she chooses it despite her previous unwillingness to do so. In this situation, too, she will suffer wrongful death.

So far in this chapter I have argued, first, that the three types of weak character that I defended in chapter three may result in patients selecting PAS and suffering wrongful death. They may die wrongfully either because PAS is not what they would have selected for themselves before the offer of PAS was made or because their choice is the

---

[9] One example of such publicity is "How to Die: Simon's Choice" (2016). This was a BBC documentary following the final months of a man who developed Motor Neurone Disease and chose to die using PAS in Switzerland.

manifestation of a vice. Second, I argued that the context of choice may also influence patients to select PAS and die wrongfully when PAS is not what they would have chosen for themselves before it was offered.

On the other side of the argument, there are patients experiencing unbearable suffering who are not influenced either by types of weak character or by the context of choice when they decide to opt for PAS, and who may thereby benefit from, rather than be harmed by, choosing it. So, in order to settle the question of whether permitting PAS would be a desirable extension of patient choice, it is necessary to weigh the harms to people who may suffer wrongful death, should PAS be permitted, against the harms to people who must endure unbearable suffering, should PAS not be permitted.

6.6 Weighing Harms in the Case of PAS

In chapter five, by drawing on some insights from two arguments about needs (Wiggins (1998) and Megone (1992)), I argued that some harms are trumping ones which should thereby always take priority over other harms. I now apply this argument to PAS in order to defend a claim that the harm to people who would suffer wrongful death, should PAS be permitted, is a trumping harm, and should always have priority over the harm to people who are suffering unbearably at the end of life and who could not relieve their suffering by dying, should PAS not be permitted. To defend this claim, I will argue that the harm of wrongful death, but not the harm of unbearable suffering, is parallel to a categorical need. Categorical needs always take priority over instrumental ones. First, however, I will argue that the harm of wrongful death and the harm of unbearable suffering are not equivalent. If, as some people claim, they are equivalent then this would defeat my argument that the harm of wrongful death trumps the harm of unbearable suffering.

6.6.1 Wrongful Death and Unbearable Suffering

One way in which it may be possible to weigh the two harms of wrongful death and unbearable suffering is to concede that they are equivalent and then to decide which should take priority using either coin-tossing after Taurek (1977), or a numbers-based decision procedure *contra* Taurek. Since some patients express a preference for death above continued existence with unbearable suffering, the harm that they face could be construed as being at least equivalent to death. So, should PAS not be legalised, then eligible patients who wish to die would suffer a harm that they claim may be equivalent

to death. If we accept this assumption, then the harms of death (but not wrongful death) and unbearable suffering may be seen to be on a par, since one of them *is* death and the other equates to death. There are two reasons, however, why this line of argument does not go through. First, I argue it is not possible to weigh the harm of unbearable suffering against death (or wrongful death) as in Taurek cases since the harms are not, in fact, equivalent. Second, even if the harm of unbearable suffering can be weighed as in Taurek cases, I have argued in chapter five that the question of whether the numbers should count in these cases has not been decisively resolved.

At first glance, an act that results in death (not wrongful death) appears worse than an act that does not result in death but results instead in a very serious harm, even when the experience expressed by some patients that their suffering is equivalent to death is taken into account. Furthermore, it may not be possible for an agent to make a valid comparison between the experience of unbearable suffering and death because the state of being dead cannot be known to her. Patients who say their suffering is *worse* than being dead cannot know this, for the same reason. However, even though being dead is unknown to an agent, she may still be harmed by death. This is because the harm in being dead (rather than in dying) is not experiential: an agent may be harmed by death because of the goods it has deprived her of. So an agent's experiences are not useful in making a comparison between unbearable suffering and being dead. Neither can the experience of suffering itself be utilised as a comparator in weighing unbearable suffering against death. The issue for patients who are suffering is the very serious one that their pain is unbearable and cannot be relieved. The comparator, however, is not a matter of suffering since there is no experience of any kind, including pain, in being dead.

Even if we accept that the harm of unbearable suffering is equivalent to death, this does not make it equivalent to the harm of *wrongful* death. Wrongful death is in a different category of harm to death, in virtue of its wrongfulness. So a fair construal of the PAS case is that both the harms are significant, but fundamentally different[10]. It is the harm of wrongful death, which may occur as a result of the harm of weak character or the harm of context of choice, which must be weighed against the harm of unbearable

---

[10] A further way in which the harms of unbearable suffering and wrongful death can be conceptualised is in terms of duties to avoid them. The positive duty to alleviate suffering is of different categorical order to the negative duty not to wrongfully kill another, which permits their survival, e.g. Foot (2005).

suffering, in order to answer whether the permitting of PAS would be a desirable extension of patient choice.

I argued in chapter five that neither Scanlon's "individualist restriction" (Scanlon, 1998) nor Kamm's irrelevant utilities (Kamm, 2005) were able to help with the weighing of different harms of the types that I am considering here. Scanlon's argument fails to help us because there are stalemate cases in which both sets of agents facing different harms may make a reasonable objection to a principle saying that the other group of people should be helped. Kamm argues that there are some harms that should give way to more significant ones, but as these are trivial ones, they also do not help us to balance the harms in the case of PAS.

A further approach to weighing harms in the case of PAS may be to draw from the literature on harms themselves. We saw in chapter five, however, that writings on harms from Hanser (2008), Thomson (2011), Shiffrin (2012), and Kahane and Savulescu (2012) do not aim to show us how these can be weighed against one another and so are of little help in the case of PAS. However, Feinberg's (1987) argument about different types of harm gave us reason to think that some harms may be prioritised and Wiggins (1998) and Megone (1992) provided insights that helped to defend a claim that some harms are trumping ones and should always take priority over other harms. So, in order to answer the question about whether or not permitting PAS is a desirable extension of patient choice, I will argue that the harm of wrongful death has a special weight that allows it to trump the harm of unbearable suffering.

6.6.2 The Harm of Wrongful Death is a Trumping Harm

My argument in chapter five on trumping harms was based on insights from arguments about different types of need. Categorical needs, following Wiggins (1998) and Megone (1992), are grounded in requirements for human survival, whilst instrumental needs are grounded in the contingent desires of agents. Megone's argument is that needs-claims may have different forces, thereby enabling them to be prioritised, and that the claims of categorical needs should always take priority over the claims of instrumental needs. Following a parallel argument that categorical-type harms can also be identified and that these should trump other types of harm, I argue that the harm of wrongful death is a categorical harm, so should trump other harms such as unbearable suffering.

The categorical needs, listed by Megone, that are necessary for human survival, include water, food, shelter and security (Megone, 1992, p.22). If these needs are not satisfied,

then it is in the nature of a human that she will die. Furthermore, if a human were wrongfully deprived of water, food etc. she would clearly die. This has an equivalent effect to wrongful death by any other means. We can see, then, that there are close similarities between the two notions of categorical needs and wrongful death. Furthermore, if categorical needs derive their force from their connection to survival, survival itself may have an even greater claim for categorical force. In the light of the similarity between categorical needs and the harm of wrongful death, then, the harm of wrongful death is a categorical harm. Since categorical harms are trumping ones, the harm of wrongful death as a result of selecting PAS is a trumping harm, and should always have priority over other harms.

The other relevant categories in this argument which seek to identify similarities between needs and harms, are instrumental needs and unbearable suffering. I now argue that these are also relevantly similar to each other. First, instrumental needs have a goal to which they are directed. One of my examples, in chapter five, of an instrumental need was rosin for a cello bow[11]. This is an instrumental need because it is necessary for a further goal, namely playing the cello. Unbearable suffering seems parallel to this need as it may harm a patient because it prevents her from achieving her goals. She may, for example, wish to go out to meet her friends but be unable to do so because she is in too much pain. Second, unbearable suffering does not appear parallel to a categorical need: it is possible (although very distressing) for a human to survive with unbearable suffering. So there are relevant similarities between instrumental needs and harms such as unbearable suffering.

6.6.2.1 Unbearable Suffering is Relevantly Distinct From Wrongful Death

It can be objected that unbearable suffering may be sufficiently severe for it to prevent patients from surviving as autonomous beings. This appears to be the view of Levinas, who claims that significant suffering may reduce an agent to a state of "supreme irresponsibility, into infancy" (Levinas, 1987, p.72). If it did so, this would give the harm of unbearable suffering a substantial weight that could be on a par with wrongful death. Furthermore, Megone's list of requirements for human survival includes "certain psychological requisites such as love" (Megone, 1992, p.22), and it is possible that a patient who is suffering unbearably may not be capable of fulfilling these needs. This

[11] This need may look at first glance as if it is a trivial one; but consider a professional cellist who would be unable to play without rosin (or a suitable alternative), and who would thereby be unable to provide for herself.

would also place patients who are suffering unbearably on the categorical harm side, alongside patients who die wrongfully.

In order to defend the claim that there is a relevant distinction between wrongful death and unbearable suffering, I argue that suffering itself is a complex psychological state and, as such, it does not foreclose the possibility of either autonomous functioning or achieving psychological requisites, especially in the context of requests for PAS. First, it is a necessary requirement of requests for PAS that patients have the capacity to make this decision, and this entails that they are autonomous, at least in this respect. Second, I argue that the necessary psychological requisites for human survival are "personal relationships" of any type (Blum, 2008), and patients with unbearable suffering in the context of PAS have these relationships. Importantly, personal relationships in a context of suffering normally trigger a compassionate response from the other party in the relationship, and this response, in turn, entails the possibility of benefit to the patient.

The complex psychology of suffering is illustrated by cases, first, in which suffering does not necessarily track physical sensations, such as pain. As an example of this, consider a woman who experiences severe pain in childbirth, but who does not necessarily suffer from this pain. Furthermore, Cassel (1982) cites examples of patients with pain such as sciatica and pain in terminal illness who, on identification of the cause of the pain, may experience a reduction in suffering. So the woman in childbirth may not suffer because of the positive context of her painful experience, and Cassel's patients undergo a reduction in suffering as a result of a psychological change mediated by increased understanding of their pain. Second, suffering can be *purely* psychological. "Existential suffering" may be experienced by patients in the absence of any physical sensations. There are many different theories about existential suffering, but common themes that emerge are threats to the "intactness of the person as a complex social and psychological entity" (Cassel, 1982), an alienating mood (Svenaeus, 2014) and an apparent lack of meaning in life (Frankl, 1997). So far, however, this account of suffering does not explain how patients who are suffering unbearably may still function autonomously and meet psychological needs for survival, thereby entailing a clear distinction from wrongful death.

It should be noted at this point that I am seeking to defend a distinction between the harm to patients ensuing from the failure to alleviate unbearable suffering *if PAS is not made a live option for them*, and the harm ensuing from death chosen by patients who have certain types of weak character *if PAS is made a live option for them*. If PAS is

made a live option for patients with unbearable suffering, then in order potentially to benefit from it, they are required to have the capacity to choose between PAS and alternative treatment options. Patients who have this capacity and then choose PAS are thereby acting autonomously, in line with their beliefs and values and despite their unbearable suffering. So, in the relevant context, namely one of being eligible for PAS, unbearable suffering does not preclude acting autonomously. Furthermore, this argument is not inconsistent with Levinas' claim (above) about the consequences of significant suffering, since these consequences may be intermittent.

A second way in which unbearable suffering may be on a par with wrongful death is if it renders a patient unable to experience "psychological requisites" that "a good human being" "probably" needs "given the kind of thing a human being is" (Megone, 1992, p.22). So Megone is not certain that humans have these requisites, but even if they are a necessity, I claim that they can be met by patients who are suffering unbearably. My first move is to say more about what psychological requisites for survival may consist in—namely types of "personal relationships" (Blum, 2008). Second, I argue that unbearable suffering is not necessarily a barrier to experiencing personal relationships, and these relationships hold the potential of benefit to the patient, for example through compassion.

Blum (2008) divides personal relationships into "categorial" and "quality" types. The categorial type describes the relationship according to societal labels including, but not limited to, familial or institutional ones, such as parent, sibling or partner/spouse, and nurse, social worker or carer. Quality types of personal relationships consist in features according to which we ascribe different values to those relationships. Blum's list of qualities includes "deep concern, involvement, commitment, care, loyalty [and] intimacy" (Blum, 2008, p.512). Clearly, these two types overlap, and some categorial relationships may or may not have certain qualities—caring, for example, is expected (but not guaranteed) of nurses and other health professionals or parents. Importantly, most people have personal relationships of one of Blum's types or another, through various types of human interaction.

If we now turn to patients who are suffering unbearably in the context of having the option of PAS, it can be seen that they, too, have at least the opportunity of personal relationships of various types. They may have family and other relationships and will be in personal relationships with care professionals engaged in the process of deciding about PAS (which is the relevant context). So the possibility of personal relationships in

this context is not foreclosed by the patient's unbearable suffering, although some patients may feel that the severity of their suffering prevents them from engaging fully in these relationships.

However, an important and potentially beneficial aspect of any relationship that patients have is the other agents' response to their suffering. In particular, patients who are suffering unbearably may expect to be shown compassion[12]. Compassion is an important response in the context of suffering because it should motivate agents to offer help to the patient. Help could be in the form of emotional or practical support, in the case of family members or friends, or could be more specific and technical, in the case of care professionals. In either case, suffering patients may benefit from a compassionate response to their suffering. Clearly, patients may not necessarily benefit from being helped by an agent (of any type) who is compassionate towards them. However, personal relationships and a compassionate response confer the possibility of benefit to patients who are suffering unbearably. Furthermore, types of therapeutic help are continuously extending their reach, e.g. the use of psilocybin in depression and anxiety in terminal illness (McCorvy et al., 2016). So, unbearable suffering in the context of PAS is relevantly distinct from wrongful death; it does not entail loss either of autonomous functioning or the psychological requisites for human survival.

Now that we have linked wrongful death and categorical needs, and unbearable suffering and instrumental needs, we can use these links to defend a claim that the harm of wrongful death trumps the harm of unbearable suffering. The force of categorical needs is what gives them precedence over instrumental needs. Humans cannot survive if their categorical needs are not satisfied whereas humans cannot satisfy their goals (except survival) if their instrumental needs are not satisfied. Furthermore, humans have to meet their categorical needs in order to meet their instrumental ones. So the harm of wrongful death, being parallel to categorical need, trumps the harm of unbearable suffering which is parallel to an instrumental need.

---

[12] Schopenhauer (1998), has argued for a foundational role in ethics for compassion. More recently, compassion's place in healthcare has been defended, e.g. Gelhaus (2012), and its absence attacked, e.g. in the Francis Report on the Mid Staffordshire Hospitals Foundation Trust Public Inquiry (2013).

6.7 Conclusion

In this chapter, I have applied the arguments in chapters one to five to the case where choice for selected patients is extended as a result of PAS being offered as an additional option. Patients are more likely to select PAS and wrongfully die if they have any of the three types of weak character that I have defended, or as a result of being adversely influenced by the context of choice.

Should patient choice not be extended by offering PAS, then certain patients with unbearable suffering may be harmed because they cannot benefit from PAS. These are patients, first, whose selection of PAS does not reflect weak character of the types I have defended, and second, who are not adversely affected by the context of choice. So, in order to decide whether patient choice should be extended by offering PAS, the harm to this group of patients must be balanced against the harm to patients who die wrongfully, should PAS be made permissible.

I argued that the harm to patients who suffer wrongful death is a trumping harm since it is parallel to a categorical need—a need whose force is grounded in facts about human nature and survival. So the harm of wrongful death should always take priority over other harms. Furthermore, the harm to patients who cannot benefit from PAS, should it not be made permissible, who claim that their life of suffering is equivalent to or worse than death, is not on a par with the harm of wrongful death and so is trumped by this harm. Drawing on all the arguments above, I conclude in this chapter that the permitting of PAS is not a desirable extension of patient choice.

Conclusions

In this dissertation, I have argued that permitting Physician Assisted Suicide (PAS) is not a desirable extension of patient choice. Should PAS be made permissible, it would become an additional option for certain people who are dying, and as we saw in chapter one, agents may be harmed in various ways if they are presented with an additional option. I argued in chapter one that there are two aspects of choice situations, namely the character of the agent and the context of choice, that may result in harm to an agent should she be offered an additional option. Furthermore, types of weak character and adverse features of the context of choice may result in harms to an agent that are different to the harms from additional options identified by Dworkin, G. (1988) and Velleman (2015). However, the harm of weak character and the harm of context of choice may additionally occur in cases used by Dworkin and Velleman to illustrate their harms.

In order to develop my argument, it was necessary in chapter two to provide an account of character. First, I defended an Aristotelian account of character as a psychological disposition of agents, in respect of their emotions and beliefs, which incorporates the capacity to make deliberated choices and to act in various ways that become more stable through habituation. Importantly, this account includes a taxonomy of character types, including vices and weak-willed, strong-willed and virtuous character states, that depend for their interpretation on respective configurations of emotion and desire, belief, choice and action. Second, in order to strengthen its plausibility, I provided a defence of this Aristotelian account against more modern ones, including those of Kant (1996), and contemporary accounts from Kupperman (1991), Goldie (2004), Nagel (1979) and Williams (1981). My defence of Aristotle against Kant was important because Kant's account has a secondary role for desire/emotion, and desire/emotion is fundamental for Aristotle in explaining his different character types, including their normative implications. I suggested that selected contemporary accounts of character do not add significantly to Aristotle's account, except in the area of constitutive psychological features of agents that are the substrate for character formation. Third, I defended Aristotle's account of character from objections that have been raised from the perspective of situationist ethics, which claims that it is situational features alone that settle how an agent behaves. I claimed that there were flaws in the situationist experiments and that an Aristotelian account of character can explain an agent's behaviour in these experimental settings. Last, I defended a central, theoretical role for

the concept of character in an account of good ethical judgement. This is important to my thesis since it helps in assessing and understanding an agent's acts (such as the selection of PAS), in ways that are more sophisticated than those allowed by moral theories, based solely on notions of right and good.

In chapter three, I utilised the account of character from chapter two and defended three types of weak character that may render an agent more susceptible to harms from being given an additional option. The three types of weak character are weakness of will (*acrasia*), undue self-deprecation—a vice which is the defect of character associated with the virtue of humility—and undue lack of confidence in one's judgements. *Acratic* behaviour was seen to be more likely, first, in situations that arouse strong emotions, and second, in novel situations. Undue self-deprecation is relevant in situations where an agent is being offered options that she may need. Last, undue lack of confidence in one's judgements may result in an agent being harmed, first, as a result of acting in line with her judgements but having undue lack of confidence in them, or second, as a result of letting go of her values as a result of lack of confidence. Any of these three types of weak character underpins the harm of weak character that an agent may suffer when she is offered an additional option.

Alongside types of weak character, the second harm I defended, that may occur when an agent is presented with an additional option, is what I term the harm of context of choice. So, in chapter four, in order to defend the harm of context of choice, I first turned to the etymology of 'context' to identify features of a choice situation that are relevant for my purposes: these are features of a choice situation that are woven into an agent's decision-making. Second, I used resources from the literatures on bounded rationality and adaptive preferences to delimit the context of choice to features of a choice situation that are either salient to the agent or should or should not be salient to her. The concepts of bounded rationality and adaptive preferences were relevant for my purposes since they are both examples of situations where features of the situation influence an agent to undergo a psychological change—a change in her preferences. This change in her preferences alters the salience to her of features of the context of choice. Last in this chapter, I used a medical analogy to draw out the normative aspects of the context of choice. These normative features explain how an agent may be harmfully influenced in a choice situation. An agent is harmed if she either finds salient features of the choice situation that she should not find salient, or does not find salient

features of the choice situation that she should find salient, in order for her to make a rational choice.

An agent may also *benefit* from having an additional option, so she may be harmed if she does not have this additional option. If a decision is being made whether or not to offer an additional option, the harms, should it be offered, have to be weighed against the harms, should it not be offered, in order to decide which side should be given priority. So, next, in chapter five, I defended a claim that some harms are trumping ones, which by their nature should always take priority over other harms. The line of my argument about trumping harms derived, first, from arguments about Taurek (1977) cases. Taurek's arguments indicate that lethal harms may take priority, but the arguments about suitable decision-procedures for Taurek cases have not been decisively resolved. Selected writings on harms, since they do not set out to weigh them, also did not allow us to infer that there are some harms that are trumping ones. However, using insights from an argument about needs (Megone 1992, Wiggins 1998) and, less so, rights as trumps (Dworkin, R., 1989), I am able to defend a claim that harms that are parallel to categorical needs are trumping ones and should always take priority over other harms.

Applying the arguments in chapters one to five to the case of PAS, I defended, in chapter six, claims that the harm of weak character of the types I described and the harm of context of choice are relevant in a situation where PAS is made permissible by being legalised. Should PAS be made permissible, certain patients may choose PAS and be harmed by suffering wrongful death if, first, this is not what they would have chosen before it was made permissible, or second, their choice is the manifestation of a vice of character.

Should PAS not be made permissible, however, certain patients who are suffering unbearably at the end of their lives, and whose suffering is unrelievable, may be harmed because they do not have the option of ending their suffering by dying. So, in order to decide whether PAS should be made permissible, the harms suffered by certain patients, should it not be made permissible, have to be weighed against the harm of wrongful death suffered by certain patients, should it be made permissible. I argued, in chapter six, that the harm to people who suffer wrongful death, as a result of selecting PAS, is a trumping harm. Furthermore, patients who are suffering unbearably in the context of PAS are capable of both autonomous functioning and meeting the psychological

requisites for human survival; in other words, the harm to them, should PAS not be made permissible, is not on a par with wrongful death. So, the harm to people who suffer wrongful death as a result of selecting PAS trumps the harm to people who are suffering unbearably at the end of life and who are denied PAS, should it not be made permissible. I conclude, then, that the permitting of PAS is not a desirable extension of patient choice.

List of References

Achtenberg, D. 2002. *Cognition of Value in Aristotle's Ethics: Promise of Enrichment, Threat of Destruction.* Albany: State University of New York Press.

Annas, J. 1993. *The Morality of Happiness.* New York: Oxford University Press.

Annas, J. 2005. Comments on John Doris's Lack of Character. *Philosophy and Phenomenological Research.* 71(3), pp.636-642.

Annas, J. 2015. Applying Virtue to Ethics. *Journal of Applied Philosophy.* 32(1), pp.1-14.

Anscombe, G.E.M. 1958. Modern Moral Philosophy. *Philosophy.* 33(124), pp.1-19.

Aristotle. 1992. *Aristotle's Eudemion Ethics Books I, II, and VIII.* Trans. Woods, M., 2nd ed. Oxford: Clarendon Press.

Aristotle. 2002. *Nichomachean Ethics.* Trans. Rowe, C., Oxford: Oxford University Press.

Athanassoulis, N. 2000. A response to Harman: virtue ethics and character traits. *The Aristotelian Society.* 100(2000), pp.215-221.

Athanassoulis, N. 2005. *Morality, moral luck, and responsibility: fortune's web.* Basingstoke: Palgrave Macmillan.

Barnes, E. 2009. Disability and Adaptive Preferences. *Philosophical Perspectives.* 23(1), pp.1-22.

Baron, M. 1997. Kantian Ethics. *Three methods of ethics: a debate.* Malden, Mass: Blackwell, pp.3-91.

Barry, B. 1965. *Political Argument.* London: Routledge and Kegan Paul.

Battin, M.P., van der Heide, A., Ganzini, L., van der Wal, G. and Onwuteaka-Philipsen, B.D. 2007. Legal physician-assisted dying in Oregon and the Netherlands: evidence concerning the impact on patients in "vulnerable" groups. *Journal of Medical Ethics.* 33(10), pp.591-597.

Baxley, A.M. 2003. Does Kantian Virtue Amount to More than Continence? *The Review of Metaphysics.* 56(3), pp.559-586.

Beiser, F. 2007. A Lament. In: Kerry, P. ed. *Freidrich Schiller, Playwright, Poet, Philosopher, Historian.* Oxford: Peter Lang, pp.233-250.

Berges, S. 2011. Why Women Hug their Chains: Wollstonecraft and Adaptive Preferences. *Utilitas.* 23(01), pp.72-87.

Blackburn, S. 2005. *The Oxford dictionary of philosophy.* Oxford: Oxford University Press.

Blum, L.A. 2008. Personal Relationships. In: Frey, R.G. and Wellman, C.H. eds. *A Companion to Applied Ethics.* John Wiley & Sons, pp.512-524.

Bovens, L. 1992. Sour Grapes and Character Planning. *Journal of Philosophy.* 89(2), pp.57-78.

Broadie, S. 2002. Commentary on Aristotle's Nicomachean Ethics. *Aristotle Nicomachean Ethics.* Oxford: Oxford University Press, pp.261-452.

Brock, D.W. 1992. Voluntary Active Euthanasia. *Hastings Center Report.* 22(2), pp.10-22.

Brown, L. 1997. What is "the mean relative to us" in Aristotle's Ethics? *Phronesis.* 42(1), pp.77-93.

Bruckner, D.W. 2009. In defense of adaptive preferences. *Philosophical Studies.* 142(3), pp.307-324.

Buchanan, A.E. and Brock, D.W. 1990. *Deciding for others: the ethics of surrogate decision making.* Cambridge: Cambridge University Press.

Cassel, E. 1982. The Nature of Suffering and the Goals of Medicine. *New England Journal of Medicine.* 306(11), pp.639-645.

Charles, D. 1984. *Aristotle's Philosophy of Action.* London: Duckworth.

Coggon, J. 2006. Arguing about physician-assisted suicide: a response to Steinbock. *Journal of medical ethics.* 32(6), pp.339-341.

Cooper, J. 1996. An Aristotelian Theory of the Emotions. In: Rorty, A.O. ed. *Essays on Aristotle's Rhetoric.* Berkley: University of California Press, pp.238-257.

Cordner, C. 1994. Aristotelian Virtue and its Limitations. *Philosophy.* 69(269), pp.291-316.

Crisp, R. 2006. Aristotle on Greatness of Soul. In: Kraut, R. ed. *The Blackwell Guide to Aristotle's Nicomachean Ethics.* Oxford: Blackwell.

Cronqvist, H. and Thaler, R.H. 2004. Design Choices in Privatized Social-Security Systems: Learning from the Swedish Experience. *The American Economic Review.* 94(2), pp.424-428.

Darley, J.M. and Batson, C.D. 1973. "From Jerusalem to Jericho": A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology.* 27(1), pp.100-108.

How to Die: Simon's Choice. 2016. BBC. 10 February, 21:00.

Director_of_Public_Prosecutions. 2010. *Policy for Prosecutors in Respect of Cases of Encouraging or Assisting Suicide.* [Online]. [Accessed 01/01/17]. Available from: http://www.cps.gov.uk/publications/prosecution/assisted_suicide_policy.pdf

Doris, J.M. 2002. Persons, situations, and virtue ethics. *Nous.* 32(4), pp.504-530.

Dworkin, G. 1988. Is more choice better than less? *The theory and practice of autonomy.* Cambridge: Cambridge University Press.

Dworkin, G. 2009. Physician Assisted Death: The State of the Debate. In: Steinbock, B. ed. *The Oxford Handbook of Bioethics.* Oxford: Oxford University Press, pp.375-392.

Dworkin, R. 1989. Rights as Trumps. In: Waldron, J. ed. Oxford: Oxford University Press, pp.153-167.

Dworkin, R. 2011. *Justice for hedgehogs.* Cambridge, Mass; London: Belknap.

Dworkin, R., Nagel, T., Nozick, R., Rawls, J., Scanlon, T.M. and Thomson, J.J. 1997. Assisted suicide: the philosophers' brief. *The New York review of books.* 44(5), pp.41-47.

Elster, J. 1983. *Sour grapes: studies in the subversion of rationality.* Cambridge: Cambridge University Press.

Elster, J. 1989. Social Norms and Economic Theory. *Journal of Economic Perspectives.* 3(4), pp.99-117.

Feinberg, J. 1987. *The Moral Limits of the Criminal Law v1: Harm to Others.* Oxford: Oxford University Press.

Flanagan, O.J. 1991. *Varieties of moral personality: ethics and psychological realism.* Cambridge, Mass; London: Harvard University Press.

Foot, P. 2005. The Problem of Abortion and the Doctrine of Double Effect. In: Steinbock, B. and Norcross, A. eds. *Killing and Letting Die.* New York: Fordham University Press, pp.266-279.

Francis, R. 2013. *Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry Volume 1: Analysis of evidence and lessons learned (part 1).* HC 898-I, London: The Stationary Office.

Frankl, V.E. 1997. *Man's Search for Meaning.* New York: Pocket Books.

Gelhaus, P. 2012. The desired moral attitude of the physician: (II) compassion. *Med Health Care and Philos.* 15, pp.397-410.

Gigerenzer, G. and Goldstein, D.G. 1996. Reasoning the fast and frugal way: models of bounded rationality. *Psychological review.* 103(4), pp.650-669.

Goldie, P. 2004. *On Personality.* London: Routledge.

Gottlieb, P. 2009. *The virtue of Aristotle's ethics*. Cambridge: Cambridge University Press.

Grüne-Yanoff, T. 2007. Bounded Rationality. *Philosophy Compass.* 2(10), pp.534-563.

Hanser, M. 2008. The Metaphysics of Harm. *Philosophy and Phenomenological Research.* LXXVII(2), pp.421-450.

Hardin, R. 2005. Trust. In: Honderich, T. ed. *The Oxford companion to philosophy.* Oxford: Oxford University Press, p.926.

Harman, E. 2009. "I'll be glad I did it" reasoning and the significance of future desires. *Philosophical Perspectives.* 23(1), pp.177-199.

Harman, G. 1999. Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error. *Proceedings of the Aristotelian society.* 99, pp.315-331.

Harman, G. 2000. The Nonexistence of Character Traits. *Proceedings of the Aristotelian Society.* 100, pp.223-226.

Hartshorne, H. and May, M. 1928. *Studies in the nature of character, vol 1, Studies in deceit.* New York: Macmillan.

Hausman, D.M. and Welch, B. 2010. Debate: To Nudge or Not to Nudge. *Journal of Political Philosophy.* 18(1), pp.123-136.

Hirose, I. 2007. Aggregation and Non-Utilitarian Moral Theories. *Journal of Moral Philosophy.* 4(2), pp.273-284.

Hope, T. 2004. Why Undervaluing Statistical People Costs Lives. *Medical Ethics: A Very Short Introduction.* Oxford: Oxford University Press, pp.26-41.

Hurka, T. 1987. Why Value Autonomy? *Social Theory and Practice.* 13, pp.361-382.

Hursthouse, R. 1988. Moral Habituation. *Oxford Studues in Ancient Philosophy.* 6, pp.210-219.

Hursthouse, R. 1999. *On Virtue Ethics.* Oxford: Oxford University Press.

Hursthouse, R. 2006a. Are Virtues the Starting Point for Moral Theory? In: Dreier, J. ed. *Contemporary Debates in Moral Theory.* Oxford: Oxford University Press, pp.99-112.

Hursthouse, R. 2006b. The Central Doctrine of the Mean. In: Kraut, R. ed. *The Blackwell Guide to Aristotle's Nicomachean Ethics* Oxford: Blackwell, pp.96-115.

Hursthouse, R. and Pettigrove, G. 2016. *Virtue Ethics.* The Stanford Encyclopedia of Philosophy, Zalta, E. ed, [Online]. [Accessed 04/01/17]. Available from: https://plato.stanford.edu/entries/ethics-virtue/

Isen, A.M. and Levin, P.F. 1972. Effect of feeling good on helping: cookies and kindness. *Journal of personality and social psychology.* 21(3), p384.

Iyengar, S. 2011. *The art of choosing.* London: Abacus.

James, W. 1981. *The Principles of Psychology.* Cambridge MA: Harvard University Press.

John, O. and Srivastava, S. 2010. The Big Five Trait Taxonomy: History, Measurement and Theoretical Perspectives. In: John, O., et al. eds. *Handbook of personality: theory and research.* New York; London: Guilford Press, pp.102-138.

Kahane, G. and Savulescu, J. 2012. The concept of harm and the significance of normality. *Journal of Applied Philosophy.* 29(4), pp.318-332.

Kahneman, D. 2012. *Thinking, fast and slow.* London: Penguin.

Kahneman, D. and Tversky, A. 1973. On the psychology of prediction. *Psychological Review.* 80(4), pp.237-251.

Kamisar, Y. 1958. Some Non-Religious Views Against Proposed "Mercy Killing" Legislation. *Minnesota Law Review.* 42(6), pp.969-1042.

Kamm, F.M. 1993. *Morality, mortality.* Oxford; New York: Oxford University Press.

Kamm, F.M. 2005. Aggregation and Two Moral Methods. *Utilitas.* 17(1), pp.1-23.

Kamtekar, R. 2004. Situationism and Virtue Ethics on the Content of Our Character. *Ethics.* 114(April), pp.458-491.

Kant, I. 1996. *The Metaphysics of Morals.* Trans. Gregor, M.J., Cambridge: Cambridge University Press.

Kant, I. 1998. *Groundwork of the metaphysics of morals.* Trans. Gregor, M.J., New York; Cambridge, U.K: Cambridge University Press.

Kawall, J. 2009. In Defense of the Primacy of the Virtues. *Journal of Ethics and Social Philosophy.* 3(2), pp.1-21.

Kaysen, S. 2000. *Girl, interrupted.* London: Virago.

Konstan, D. 2016. *Epicurus.* The Stanford Encyclopedia of Philosophy, Zalta, E. ed, [Online]. [Accessed 26/12/16]. Available from: https://plato.stanford.edu/entries/epicurus/

Kristjánsson, K. 2008. An Aristotelian Critique of Situationism. *Philosophy.* 83(01), pp.55-76.

Kupfer, J. 2003. Perspective of Humility. *Pacific Philosophical Quarterly.* 84, pp.249-269.

Kupperman, J.J. 1991. *Character.* Oxford: Oxford University Press.

Lang, G. 2005. Fairness in Life and Death Cases. *Source: Erkenntnis.* 62(3), pp.321-351.

Levinas, E. 1987. *Time and the Other: (and Additional Essays).* Trans. Cohen, R.A., Pittsburgh: Duquesne University Press.

Levine, R. 1988. *The Ethics and Regulation of Clinical Research.* New Haven: Yale University Press.

Locke, J. 1979. *An essay concerning human understanding.* Oxford: Clarendon Press.

Lockwood, M. 1988. Quality of Life and Resource Allocation. *Royal Institute of Philosophy Lecture Series.* 23(1988), pp.33-55.

Louden, R.B. 1984. On some vices of virtue ethics. *American Philosophical Quarterly.* 21(3), pp.227-236.

Louden, R.B. 1986. Kant's Virtue Ethics. *Philosophy.* 61(238), pp.473-489.

Lübbe, W. 2008. Taurek's No Worse Claim. *Philosophy & Public Affairs.* 36(1), pp.69-85.

Manson, N. and O'Neill, O. 2008. *Rethinking Informed Consent in Bioethics.* Cambridge: Cambridge University Press.

McCorvy, J.D., Olsen, R.H.D. and Roth, B.L. 2016. Psilocybin for Depression and Anxiety

Associated with Life-threatening Illnesses. *Journal of Psychopharmacology.* 30, pp.1209-1210.

Megone, C.B. 1992. What is Need? In: Corden, A., et al. eds. *Meeting Needs in an Affluent Society: A Multi-disciplinary Perspective.* Aldershot: Avebury, pp.12-30.

Megone, C.B. 1998. Aristotelian ethics. In: Chadwick, R.F. ed. *Encyclopedia of applied ethics.* San Diego: Academic Press, pp.209-232.

Mele, A.R. 1981. Choice and Virtue in the Nicomachean Ethics. *Journal of the History of Philosophy.* 4(19), pp.405-423.

Meyer, L. 2016. *Intergenerational Justice*. The Stanford Encyclopedia of Philosophy, Zalta, E. ed, [Online]. [Accessed 25/12/16]. Available from: https://plato.stanford.edu/entries/justice-intergenerational/

Milgram, S. 1963. Behavioral study of obedience. *The Journal of Abnormal and Social Psychology*. 67(4), pp.371-378.

Mill, J.S. 1982. *On liberty.* Harmondsworth: Penguin.

Miller, C. 2003. Social Psychology and Virtue Ethics. *The Journal of Ethics.* 7(4), pp.365-392.

Munzel, G.F. 1999. *Kant's conception of moral character: the "critical" link of morality, anthropology, and reflective judgment.* Chicago, Ill: University of Chicago Press.

Nagel, T. 2013. Death. *Mortal questions.* Cambridge; New York: Cambridge University Press, pp.1-10.

Newell, A. and Simon, H. 1973. Human Problem Solving. *Contemporary Sociology*. 2(2), pp.169-169.

Nussbaum, M.C. 1986. *The fragility of goodness: luck and ethics in Greek tragedy and philosophy.* Cambridge: Cambridge University Press.

Nussbaum, M.C. 2001. Symposium on Amartya Sen's Philosophy : 5 Adaptive Preferences and Women's Options. *Economics and Philosophy.* 17, pp.67-88.

O'Neill, O. 2008. Questions of life and death. *Lancet.* 372(9646), pp.1291-1292.

Parfit, D. 1984. *Reasons and persons.* Oxford: Clarendon Press.

Price, A. 2006. Acrasia and Self-control. In: Kraut, R. ed. *The Blackwell Guide to Aristotle's Nicomachean Ethics.* Oxford: Blackwell, pp.234-254.

Prinz, J. 2009. The normativity challenge: Cultural psychology provides the real threat to virtue ethics. *The Journal of Ethics.* 13(2), pp.117-144.

Purshouse, C. 2015. A Defence of the Counterfactual Account of Harm. *Bioethics.* I, pp.1-9.

Reeve, A. 1990. Individual Choice and the Retreat from Utilitarianism. In: Allison, L. ed. *The Utilitarian Response.* London: Sage, pp.98-119.

Richards, N. 1988. Is Humility a Virtue? *American Philosophical Quarterly.* 25(3), pp.253-253.

Sarch, A. 2008. What's Wrong With Megalopsychia? *Philosophy.* 83(02), pp.231-253.

Sauvé Meyer, S. 2006. Aristotle on the Voluntary. In: Kraut, R. ed. *The Blackwell Guide to Aristotle's Nicomachean Ethics.* Oxford: Blackwell, pp.137-157.

Scanlon, T. 1998. *What we owe to each other.* Cambridge, Mass.: Harvard University Press.

Scarre, G. 2013. The Continence of Virtue. *Philosophical Investigations.* 36(1), pp.1-19.

Schelling, T.C. 1960. *The strategy of conflict.* London: Oxford University Press.

Schopenhauer, A. 1998. *On the Basis of Morality.* Trans. Payne, E.F.J., Indianapolis, Ind: Hackett Pub.

Schwalbe, U. and Walker, P. 2001. Zermelo and the Early History of Game Theory. *Games and Economic Behavior Journal of Economic Literature Classification Numbers: B19.* 34(70), pp.123-137.

Schwitzgebel, E. 2016. *Introspection.* The Stanford Encyclopedia of Philosophy, Zalta, E. ed, [Online]. [Accessed 24/12/16]. Available from: https://plato.stanford.edu/entries/introspection/

Scitovsky, T. 1976. *The joyless economy: an inquiry into human satisfaction and consumer dissatisfaction.* New York: Oxford University Press.

Sherman, N. 1985. Character, Planning, and Choice in Aristotle. *The Review of Metaphysics.* 39(1), pp.83-106.

Sherman, N. 1989. *The fabric of character: Aristotle's theory of virtue.* Oxford: Clarendon.

Shiffrin, S.V. 2012. Harm and Its Moral Significance. *Legal Theory.* 18(03), pp.357-398.

Shleifer, A. 2012. Psychologists at the Gate: A Review of Daniel Kahneman's Thinking, Fast and Slow. *Journal of Economic Literature.* 50(4), pp.1080-1091.

Simon, H. 1957. *Models of Man.* New Jersey: Wiley.

Singer, P. 1977. Freedoms and Utilities in the Distribution of Health Care. In: Dworkin, G., et al. eds. *Markets and Morals.* Washington: Hemisphere Publishing Corporation, pp.149-174.

Steinbock, B. 2005. The case for physician assisted suicide: not (yet) proven. *Journal of medical ethics.* 31(4), pp.235-241.

Svenaeus, F. 2014. The Phenomenology of Suffering in Medicine and Bioethics. *Theoretical Medicine and Bioethics.* 35(6), pp.407-420.

Taurek, J.M. 1977. Should the numbers count? *Philosophy & public affairs.* 6(4), pp.293-316.

Thaler, R.H. and Sunstein, C.R. 2008. *Nudge: improving decisions about health, wealth, and happiness.* New Haven, Conn; London: Yale University Press.

The_Behavioural_Insights_Team. 2016. *The Behavioural Insights Team.* [Online]. [Accessed 24/12/16]. Available from: www.behaviouralinsights.co.uk

The_National_Council_for_Palliative_Care. 2016. *Dying Matters.* [Online]. [Accessed 26/12/16]. Available from: http://www.dyingmatters.org

Thomson, J.J. 2011. More On The Metaphysics of Harm. *Philosophy and Phenomenological Research.* 82(2), pp.436-458.

Tversky, A. and Kahneman, D. 1986. Rational Choice and the Framing of Decisions. *The Journal of Business.* 59(4), pp.S251-S278.

Tversky, A. and Kahneman, D. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty.* 5(4), pp.297-323.

Tversky, A. and Simonson, I. 1993. Context-dependent preferences. *Management Science.* 39(10), pp.1179-1190.

UK_Government. 2012. *The hacking of Milly Dowler's telephone.* [Online]. [Accessed 26/12/16]. Available from: http://www.publications.parliament.uk/pa/cm201012/cmselect/cmcumeds/903/90307.htm

UK_Government. 2016. *Assisted Dying (No.2) Bill 7.* [Online]. [Accessed 09/01/17]. Available from: http://services.parliament.uk/bills/2015-16/assisteddyingno2.html

Urmson, J.O. 1973. Aristotle's Doctrine of the Mean. *American Philosophical Quarterly*. 10(3), pp.223-230.

Velleman, J.D. 2003. Narrative Explanation. *Philosophical Review.* 112(1), pp.1-25.

Velleman, J.D. 2015. *Beyond Price.* Cambridge, UK: Open Book Publishers.

Vranas, P. 2005. The indeterminacy paradox: Character evaluations and human psychology. *Nous.* 35(1), pp.1-42.

Watson, G. 1990. On the Primacy of Character. In: Flanagan, O. and Rorty, A. eds. *Identity, Character and Morality*. Cambridge Mass: MIT Press, pp.449-469.

Wenar, L. 2015. *Rights.* The Stanford Encyclopedia of Philosophy, Zalta, E. ed, [Online]. [Accessed 01/01/17]. Available from: https://plato.stanford.edu/entries/rights/

Wiggins, D. 1975. Deliberation and practical reason. *Proceedings of the Aristotelian Society.* 76, pp.29-51.

Wiggins, D. 1998. *Needs, Values, Truth.* Third ed. Oxford: Clarendon Press.

Williams, B. 1973. A Critique of Utilitarianism. In: Smart, J.J.C. and Williams, B. eds. *Utilitarianism: for and against.* London: Cambridge University Press, pp.77-150.

Williams, B. 2009. Life as Narrative. *European Journal of Philosophy.* 17(2), pp.305-314.

A World Without Down's Syndrome. 2016. BBC. 5 October, 21:00.

Zermelo, E. 1913. Uber eine Anwendung der Mengenlehre auf die Theorie des Schachspiels. *Proc. Fifth Congress Mathematicians.* pp.501-504.

List of Abbreviations

AP	Adaptive Preferences
BR	Bounded Rationality
D1-7	Dworkin's seven harms from additional options
IQ	Intelligence Quotient
MRI	Magnetic Resonance Imaging
NE	Nicomachean Ethics
ODDA	Oregon Death with Dignity Act
PAS	Physician Assisted Suicide
V1-2	Velleman's two harms from additional options

Appendix: Safeguarding in Physician Assisted Suicide

Supporters of PAS claim that it is possible to build adequate safeguarding measures into any provision of PAS in order to protect vulnerable people from requesting PAS. So, according to them, it would not be possible for anyone to wrongfully die under the conditions stipulated in an Act of Parliament that legalises PAS. In order to take advantage of an offer of PAS, the patient must have "a voluntary, clear, settled and informed wish to end his or her own life" (Assisted Dying (No. 2) Bill, 2016, p.1). Furthermore, patients who have a psychiatric illness such as depression are identified by medical screening and deemed ineligible for PAS. There are, however, reasons to think that, even if a patient meets these requirements, she may still select PAS as a result of either weak character or the context of choice and thereby die wrongfully. O'Neill (2008) has said that

> "[i]f such legislation is to work, capacity for autonomous choice has to be strictly and stringently determined. [...] The draft bills proposed many safeguards, but did not deal with the reality that those for whom the legislation was intended are in situations of acute dependence on others. How are we to distinguish requests to be killed that express individual autonomy, from requests that express compliance with the (unspoken) desires of burdened carers and relatives, not to mention expectant heirs? Legislation to make assisted dying lawful needs not only to prohibit action where a request reflects momentary despair—an issue it sought to address—but to prohibit action on requests that reflect an individual's weary compliance, indeed deference, rather than their autonomy. In a world of ideal, if mythic, rational beings, whose choosing was guaranteed to be wholly autonomous, assisted dying legislation might not be risky; but that is not our world. The philosopher Bernard Williams was, I think, right in suggesting that "we should not put too much weight on the fragile structure of the voluntary" (O'Neill, 2008).

O'Neill suggests here that there may be aspects of the context, such as being acutely dependent, that may influence the patient. Furthermore, an apparently "voluntary, clear, settled and informed wish" is not inconsistent with each of the three types of weak character that I have defended. An *acratic* action may be "voluntary, clear, [and] settled", as may an action that arises from either undue self-deprecation or undue lack of confidence in judgements. Furthermore, all these actions may appear to be properly informed (O'Neill, 2008). Each of the three types of weak character are voluntary—in and of themselves they are uninfluenced by external factors other than the guidance that is implicit in the process of habituation. Similarly, a patient with a weak character may be settled in her selection of an option. Last, I argued that on one account, *acrasia* may be more likely if the agent is uninformed. However, this is not a necessary condition for *acrasia* since the *acratic* patient may act in line with her irrational emotions and against her reasoned choice, despite being fully informed. So what is established in this

appendix is that current safeguarding measures are insufficient for preventing patients from suffering wrongful death, should they be offered the option of PAS.